Marine and Environment (2023) Volume 12 Issue 1
doi: 10.18686/me.v12i1.9205

ORIGINAL RESEARCH ARTICLE

A deep learning algorithm and software for photo identification of the Indo-Pacific humpback dolphin (*Sousa chinensis*)

Fuxing Wu[1], Neng Chen[2], Weilin Gao[2], Mengda Yu[2], Siying Huang[2], Mingding Zhong[1], Zhulin An[2]*, Xianyan Wang[1]*

[1] *Third Institute of Oceanography, Ministry of Natural Resources, Daxue Road 178, Xiamen 361005, Fujian Province, People's Republic of China*

[2] *Xiamen Data Intelligence Academy of ICT, CAS, Fengqi Road 208-3, Software Park Phase 3, Xiamen 361021, Fujian Province, People's Republic of China*

*** Corresponding author:** Zhulin An, anzhulin@ict.ac.cn; Xianyan Wang, wangxianyan@tio.org.cn

ABSTRACT

Deep neural networks have been increasingly used to identify individual animals in ecological studies by learning and distinguishing their naturally occurring marks or features. Traditional individual animal recognition requires prior knowledge and experience, which can be time-consuming and inefficient. In this paper, a distinctive deep learning framework that automatically reidentifies individual Indo-Pacific humpback dolphins (*Sousa chinensis*) from photos was proposed. For most dolphin species with a dorsal fin, this feature is reliably used to identify and distinguish individuals in studies that require distinction between members of a group or population. Feature cutting and background removing strategies were added to allow a focus on local information. Knowledge distillation was also applied to improve the robustness of the framework. Additionally, an automatic dolphin recognition software suite for cetologists that may reduce the amount of effort and time required to manually confirm individual dolphin ID from photographs had been developed. In the end, the effectiveness of applying this deep neural network approach for individual Indo-Pacific humpback dolphin recognition had been demonstrated.

Keywords: Indo-Pacific humpback dolphin; computer vision; re-identification; knowledge distillation; recognition software; *Sousa chinensis*

ARTICLE INFO

Received: 5 July 2023
Accepted: 7 August 2023
Available online: 31 August 2023

1. Introduction

Many ecological studies, such as examination of wild animal population parameters, migration, habitat usage, and social behavior, involve individual reidentification[1–3]. Long-term follow up studies allow researchers to obtain some insights into wild animal population distribution and behavior, which can provide further details and information to promote their protection[4]. Traditional animal tracking methods involve implanting identification sensors into the animal or attaching sensors to their bodies, which can be invasive and may negatively impact their actions and behavior[5]. Previous studies have confirmed that the majority of animals can be identified by their natural scars and marks (i.e., exterior natural patterns or features), such as whiskers on polar bears (Ursus Maritimus) and notches, nicks, scars or pigments on or around the dorsal fin of cetaceans[6,7]. Use of natural marks to recognize and reidentify individuals is non-invasive and usually does not disturb their normal behaviors; this approach has been applied in many fields of zoology (e.g., Auger-Méthé and Whitehead[8]).

Although there are animal recognition methods that do not require implants that affect behavior, they often require a large quantity of resources and time and the quality of recognition is affected by repetitive artificial comparison. In recent years, pedestrian reidentification using deep convolutional networks has recently contributed to the success of modern surveillance systems[9]. Therefore, we adopted and transferred this technology to the recognition of individual animals, which could reduce the effort and resources required of biologists.

It is important for marine biologists to track and study the movement pattern, social behavior, reproduction and survival of dolphins, which requires individual recognition of dolphins[3,10]. The Indo-Pacific humpback dolphin (*Sousa chinensis*) is recognized by its unique spots and dorsal fin notches[11], making it an ideal species to test our animal reidentification software application. However, recognizing individual animals is different from reidentifying a pedestrian. Specifically, people in images or photos normally walk upright with no angle deflection, while the angle of dolphin dorsal fins in images varies as they sometimes jump out of the ocean, which brings challenges to feature extraction and feature alignment in deep neural networks.

In this study, an alternative way for Indo-Pacific humpback dolphin reidentification is presented. During the training phase, an object detection algorithm to localize dolphin dorsal fins in the images with bounding boxes was used[12]. Then, key points based on the outline of the dolphin dorsal fins and corrected the dorsal fin bounding boxes in the horizontal direction according to the points was identified. A small range of random angle rotation was applied to the corrected bounding boxes for data augmentation to increase the robustness. To make full use of the local information and eliminate the interference of the background information, the dorsal fin was cut into different parts and removed the background. Moreover, knowledge distillation was adopted to improve the generalization and expressivity of the network for large deflection angles[13]. In the inference stage, an object detection algorithm was used to cut out the fin bounding boxes and the postprocessed boxes was feed to the reidentification network (student network) for individual identification.

2. Methods

2.1. Pedestrian reidentification

Pedestrian reidentification in computers is used to determine whether a specific pedestrian is contained in an image or video sequence. Most pedestrian reidentification algorithms use deep learning methods based on metric learning[14], while some studies treat reidentification as a classification problem[15]. Recent research combines metric loss with classification loss to improve the expressivity of deep networks[16]. Researchers have also used local information to improve the networks' global representation ability. These local information extraction methods include lateral cutting and extraction of key points of human posture. Wang et al.[17] used horizontal slicing to extract local features. Sun et al.[18] added an attention mechanism to align the slices of different parts while dicing pedestrians. Zhao et al.[19] used key points of human posture to intercept the ROI of various parts of the body, sent each part to the network to extract features and merged the features to align the body parts. Pedestrian reidentification is also regarded as an image retrieval problem and is heavily used in the field of object and biological reidentification.

2.2. Biological identification

Reidentification technology has been applied to individual animal recognition in recent years. Liu et al.[20] applied reidentification technology to individual tiger recognition. They used tiger key points to cut out different limbs and inserted them into corresponding network branches to improve accuracy. Krschens et al.[21] used YOLO to detect the heads of elephants and fed the results into a classification network to identify individual elephants. Biologists have recently adopted different methods for the individual identification of dolphins. For instance, Bouma et al.[22] used GoogLeNet to detect dorsal fins and applied triplet loss as the loss function. However, the authors did not study the specific problem that dolphins can be tilted at a large angle within an image. Weideman et al.[23] calculated the integral curvature of the edge of the dorsal fin and used an effective matching algorithm to identify whales and dolphins. Hughes and Burghardt[24] identified great white sharks by encoding the shape of the dorsal fin. Feng et al.[25] applied deep learning technology to individual identification of Indo-Pacific humpback dolphin. They first detected the dorsal fin and then modeled the task as a common image classification problem. However, the paper regarded the task as an image classification problem and did not further explore the problem of characteristics and angles of the dorsal fin.

This work presents an end-to-end algorithm process that learns local dolphin fin information and adapts to large fin angular rotation. The dorsal fin is detected by an object detection algorithm and then affined by correcting the orientation of the dorsal fin to the horizontal direction to reduce the effect of feature nonalignment, as shown in **Figure 1**. Teacher-student dual networks learn the corrected dorsal fin with a large range of angle rotations to improve the robustness against large angles. In the inference stage, a database and separate dolphin fin images according to labeled IDs is built. Then, for unseen images, dorsal fin detection is executed, and the results are sent to the reidentification network to search for the ID in the database.

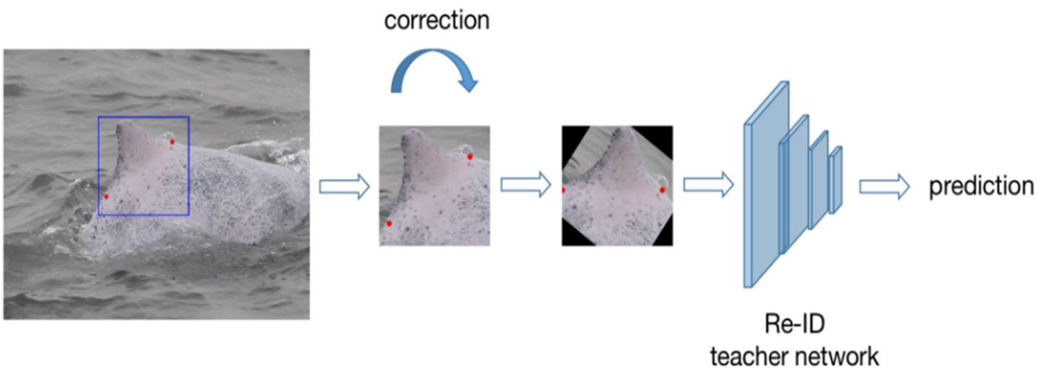

Figure 1. Individual recognition in the training stage.

2.3. Dorsal fin detection

The surface markings on the dorsal fin can be used to identify individual dolphins, making it the target of object detection. To unify the angle of dorsal fins in a number of images, two key feature points was designed, as shown in **Figure 2**. The points are defined as the intersections between the bottom line of the dorsal fin and the front and back boundaries. We find that algorithms such as face detection return inconsistent and inaccurate key point predictions because the bottom line of the dorsal fin is usually blurred[26]. Therefore, the YOLO algorithm was applied to detect dorsal fin locations and label them with bounding boxes[27]. The labeled key points were used during the training phase in order to uniformly correct dorsal fins in the horizontal direction (**Figure 1**).

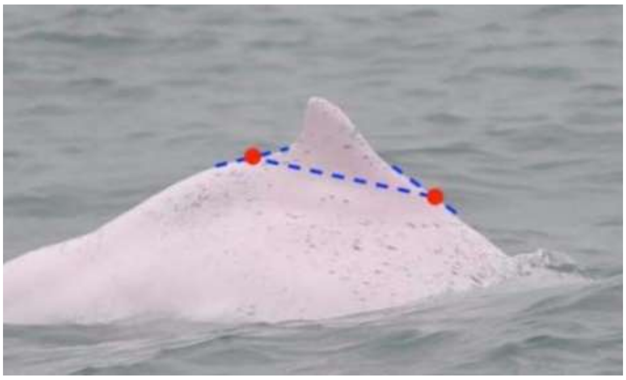

Figure 2. Key points of a dorsal fin.

2.4. Dorsal fin identification

The corrected dorsal fins were augmented by adding a random slight angle rotation to boost the generalization capability of the model. To fully extract local information and global information of dorsal fins, a three-branch structure after a convolutional backbone was adopted in the deep neural network[17]. The upper branch extracted the global features of the dorsal fins, and the middle branch taken the feature map that was horizontally cut into two parts. The upper part was further divided vertically by a ratio of 1:2:1, keeping only the middle portions and excluding the background information. The bottom branch extracted the feature map that has three equal horizontal divisions, which were the 1:1:1 vertical division upper portion, the 1:4:1 vertical division middle portion and bottom portion which was not divided. The background feature blocks were also removed to further refine the local information (**Figure 3** and **Figure 4**).

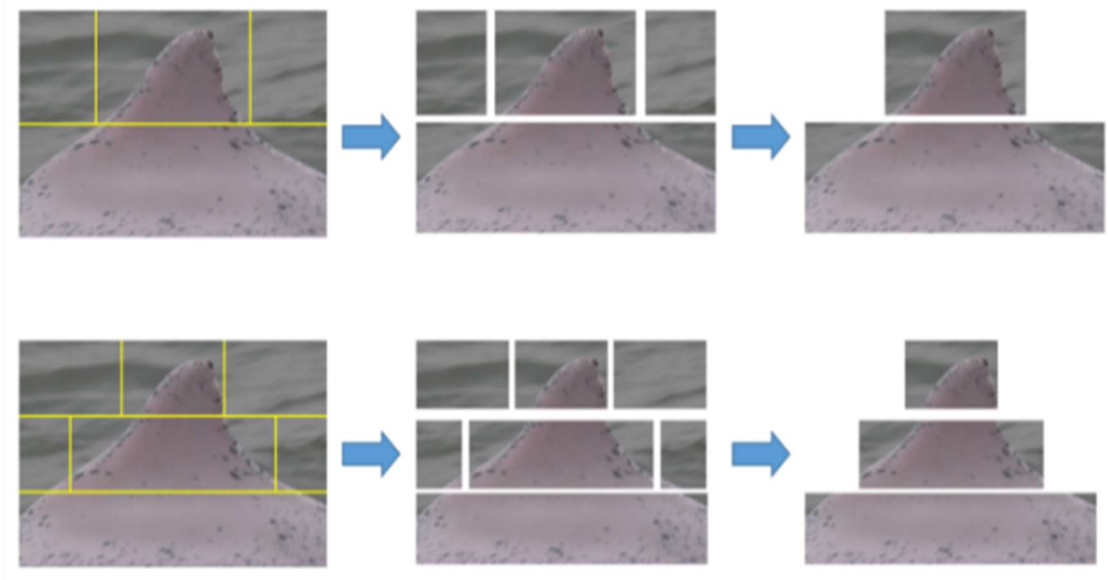

Figure 3. The original image mapped by feature cutting.

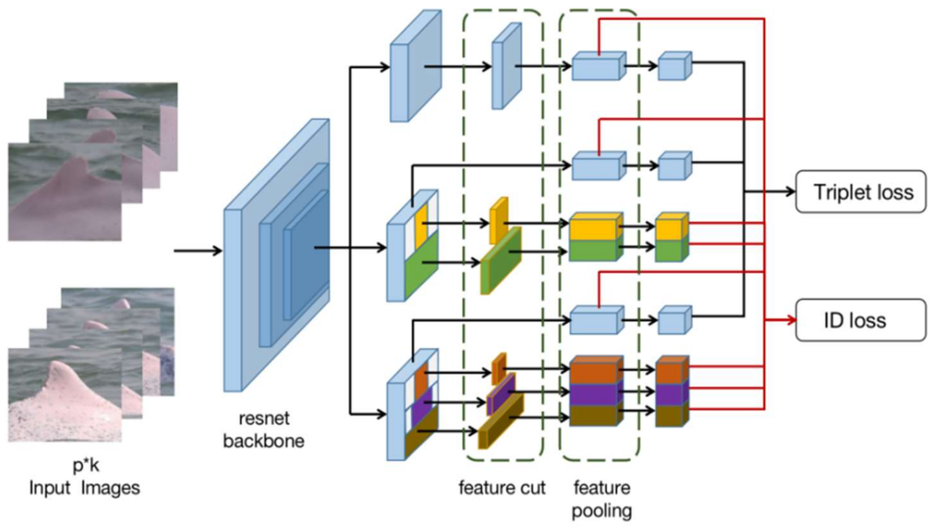

Figure 4. The reidentification network.

2.5. Knowledge distillation

When a dolphin leaps out of or dives into the water, the angle between its dorsal fin and the water surface can change significantly. We introduced a teacher-student model to learn this large angle change (**Figure 5**). The teacher model was trained on a dataset of dorsal fin images with a wide range of angles. The student model was then trained on the same dataset, but with the images randomly rotated by an angle from 25 to 40 degrees. The objective was to shrink the difference between the output features of the teacher and student models by minimizing their MSE loss so that the student model can recognize dolphins regardless of the magnitude of the angle between the dorsal fin and the water. A rotation probability hyperparameter was designed to allow the student model to memorize nonrotated images.

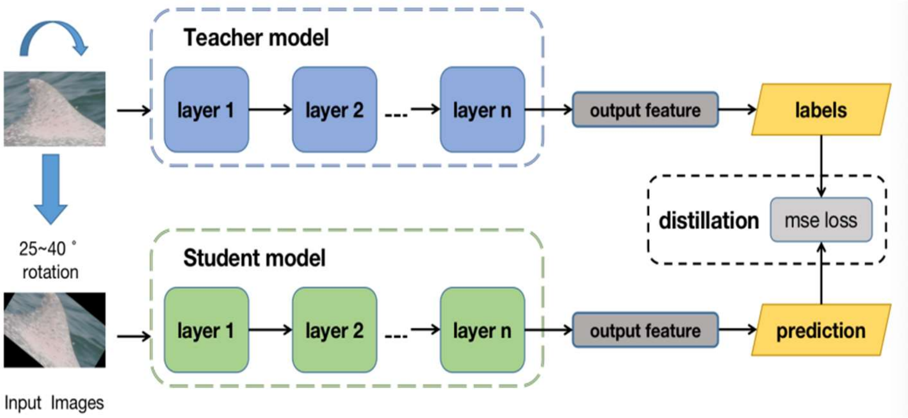

Figure 5. Teacher-student network.

2.6. Loss function

Triplet loss and classification loss was combined when training the reidentification network[28]. The output feature was used to calculate the cross-entropy loss and was then reduced to calculate the triplet loss. As shown in Equation (1), λ_1 and λ_2 are the weights of the cross-entropy loss and triplet loss, respectively: We set λ_1 to 2 and λ_2 to 1 in our experiments. In Equation (2), p is the ground truth and q is the predicted label. In Equation (3), || || denotes Euclidean distance. The superscript a represents the anchor in triplet loss, p denotes positive, n denotes negative, and α refers to the minimum interval between the distance of x^a and x^n and the distance of x^a and x^p, which we set to 4.

$$L_{total} = \lambda_1 L_{cross} + \lambda_2 L_{triple} \tag{1}$$

$$L_{cross} = -\sum_{i=1}^{n} p(x_i) \log q(x_i) \tag{2}$$

$$L_{triple} = \sum_{n=i}^{N} [\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha] \tag{3}$$

An MSE loss function was also applied in the teacher-student networks. The output feature of the teacher network was a soft label, and the output of the student network was the predicted values. In contrast to the original knowledge distillation network, independent supervision loss in the student network was removed since we focused on the model's generalization to a varying change in angle.

$$J(w,b) = \frac{1}{2n} \sum_{i=i}^{n} \|a_i - y_i\|^2 \tag{4}$$

2.7. Dataset

Our dataset consists of Indo-Pacific humpback dolphin images collected by the Third Institute of Oceanography, Ministry of Natural Resources. Due to the small population size of this species in the Xiamen Bay, our dataset has only 55 individuals, with a total of 2177 images. Moreover, the dataset was divided into a training set (90%) and a test set (10%).

2.8. Implementation

The input size of the reidentification network was 320 × 320 pixels. Several measures were taken to improve the model's generalization ability: 1) random horizontal flipping; 2) padding 10 pixels around the image and randomly cropping a 320 × 320 patch; 3) random rotation from –10 to 10 degrees; 4) randomly erasing a 16 × 16 pixels patch from an image.

In the teacher-student network, the original image was input to the teacher model, and the image rotated by a random angle from 25° to 40° with a probability of 0.7 was input to the student network.

A ResNet50 pretrained from ImageNet was applied as the backbone[29,30]. The training was optimized for 500 epochs by using SGD (Stochastic Gradient Descent method) with a batch size of 32. Each batch contained 4 IDs with 8 samples for each ID. The initial learning rate was set to 0.01 and was decreased by a factor of 0.1 at epochs 320 and 380, respectively. The object detection and reidentification models was trained on a NVIDIA 2080Ti with PyTorch implementation, while the teacher-student model was trained on a NVIDIA V100.

Dolphin recognition consisted of two components: dorsal fin detection and reidentification. the object detection algorithm was relatively sophisticated, and strategies was designed to address the fluctuation of the dorsal fin angle. In this paper, the main focus is on the reidentification results.

The algorithms were assessed by calculating the mean average precision (mAP) and Rank-k accuracy. **Table 1** shows the experiment results. Ablation experiments was conducted with two settings: 1) local information cutting and 2) knowledge distillation.

Table 1. Ablation results.

Setting			Precision				
a	**b**	**c**	**mAP**	**Rank-1**	**Rank-3**	**Rank-5**	**Rank-10**
			0.8469	0.8272	0.8346	0.8493	0.8713
√			0.8762	0.8529	0.8640	0.8860	0.9191
√	√		0.8185	0.7904	0.8456	0.8750	0.9007
√		√	0.8806	0.8640	0.8713	0.8897	0.9191

a: With feature cutting branch; b: Re-ID network with large-angle image; c: Knowledge distillation with large-angle image.

3. Results

As shown in **Table 1**, the addition of a cutting branch effectively increases the accuracy due to rich local information of the dorsal fin. The functionality of knowledge distillation was also compared. When the model was trained with images rotated by a large angle at the very beginning, the accuracy and mAP decreased. By contrast, knowledge distillation resulted in the highest Rank-1 accuracy and mAP. The results on the test dataset are shown in **Figure 6**. **Figure 7** illustrates the recognition results using the software developed based on our algorithms. The algorithms achieve a mAP of 0.8806 and Rank-1 accuracy of 0.8640. On the basis of the algorithms, we developed a dolphin reidentification software suite that was applied to practical research activities.

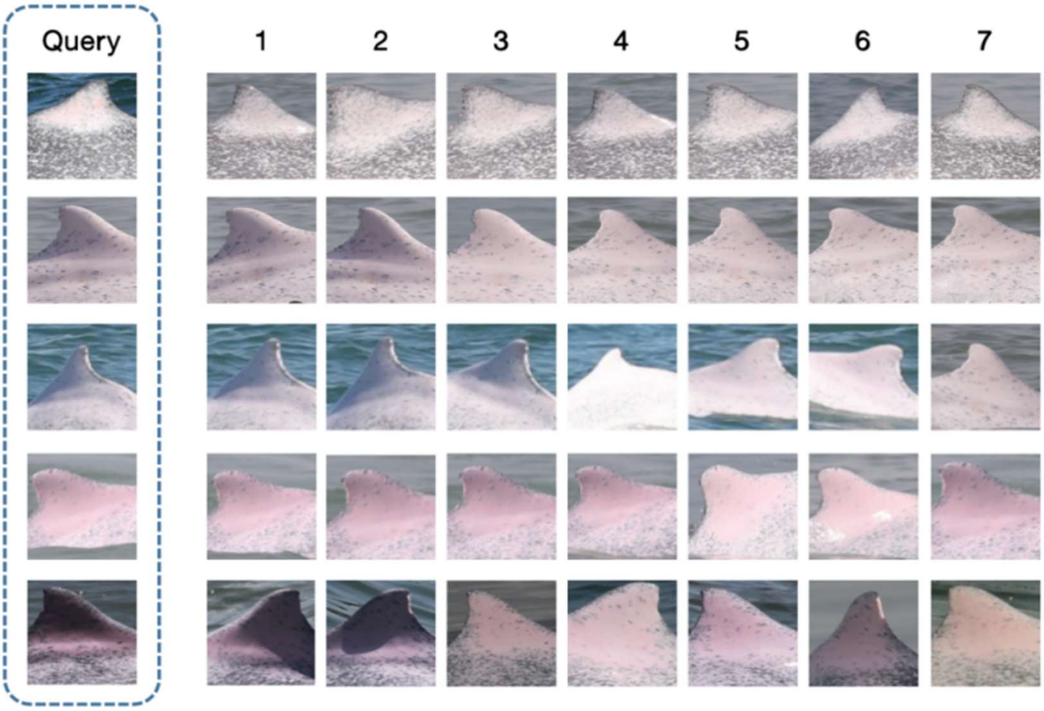

Figure 6. Retrieval results on test set.

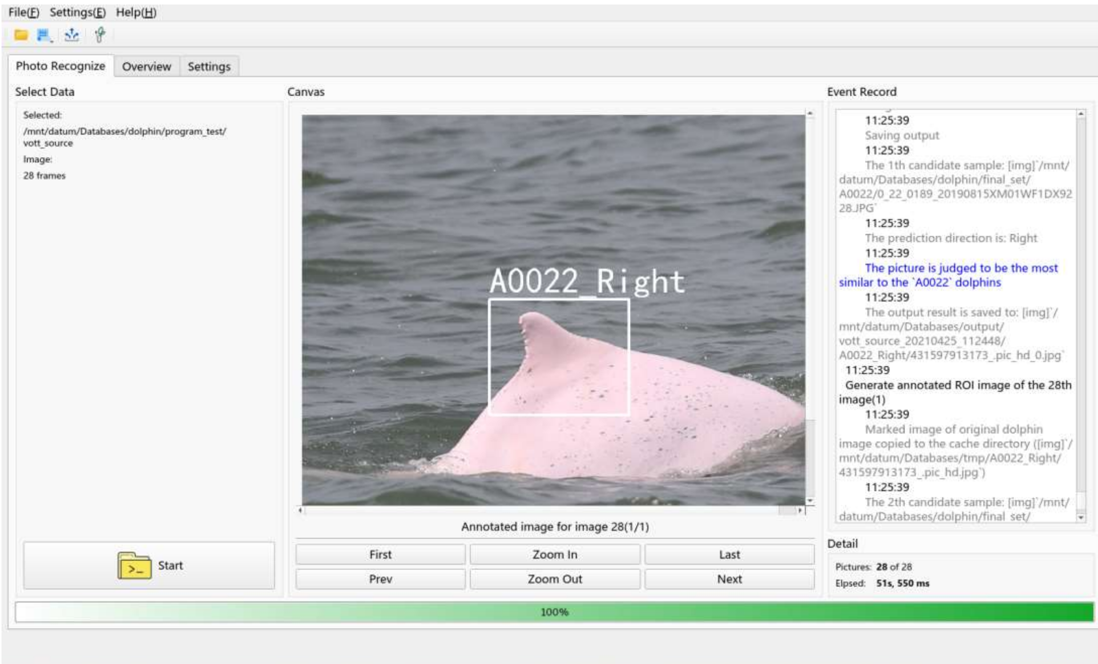

Figure 7. Recognition results.

4. Conclusions

In this paper, an end-to-end dolphin reidentification process based on deep neural networks that is capable of detecting the dorsal fins of dolphins and reidentifying individual dolphins was proposed. In addition, the proposed method in this paper can also be attempted for identification research of other species to improve work efficiency and information technology level in the future research.

The major contributions in this paper are summarized as follows:

(1) Pedestrian reidentification technology was applied to individual dolphin reidentification via deep learning and provided a unique end-to-end dolphin reidentification software suite.

(2) A method was introduced to correct a wide range of the dolphin dorsal fin angle via affine transformation to align feature maps in networks and a reidentification network was further constructed to make full use of the local dorsal fin information.

(3) Knowledge distillation was utilized to increase the network's robustness and generalization to large-angle dorsal fins.

Author contributions

Conceptualization, FW, NC, ZA, and XW; methodology, NC and WG; software, NC and WG; validation, NC, FW, ZA, and XW; formal analysis, NC and FW; investigation, FW and MZ; resources, FW and NC; data curation, FW and MZ; writing—original draft preparation, NC and FW; writing—review and editing, FW, NC, and MY; visualization, NC; supervision, FW, ZA, and XW; project administration, SH; funding acquisition, FW. All authors have read and agreed to the published version of the manuscript.

Funding

This work was financially supported by the Fujian Provincial Natural Science Foundation (Nos. 2022J05082).

Conflict of interests

The authors declare no conflict of interest.

References

1. Kreho A, Kehtarnavaz N, Araabi B, et al. Assisting manual dolphin identification by computer extraction of dorsal ratio. *Annals of Biomedical Engineering* 1999; 27: 830–838. doi: 10.1114/1.235
2. Wu F, Wang X, Ding X, et al. Distribution pattern of Indo-Pacific humpback dolphins (*Sousa chinensis*) along coastal waters of Fujian Province, China. *Aquatic Mammals* 2014; 40(4): 341–349. doi: 10.1578/AM.40.4.2014.341
3. Wang X, Wu F, Turvey ST, et al. Social organization and distribution patterns inform conservation management of a threatened Indo-Pacific humpback dolphin population. *Journal of Mammalogy* 2015; 96(5): 964–971. doi: 10.1093/jmammal/gyv097
4. Wang X, Wu F, Chang W, et al. Two separated populations of the Indo-Pacific humpback dolphin (*Sousa chinensis*) on opposite sides of the Taiwan Strait: Evidence from a larger-scale photo-identification comparison. *Marine Mammal Science* 2016; 32(1): 390–399. doi: 10.1111/mms.12257
5. Nichols, JD. Capture-recapture models: Using marked animals to study population dynamics. *BioScience* 1992; 42(2): 94–102. doi: 10.2307/2937171
6. Würsig B, Jefferson TA. Methods of photo-identification for small cetaceans. *Reports of the International Whaling Commission* 1990; 12: 43–52.
7. Hammond PS, Mizroch SA, Donovan GP. *Individual recognition of cetaceans: Use of photo-identification and other techniques to estimate population parameters*. International Whaling Commission; 1990.
8. Auger-Méthé M, Whitehead H. The use of natural markings in studies of long-finned pilot whales (Globicephala Melas). *Marine Mammal Science* 2006; 23(1): 77–93. doi: 10.1111/j.1748-7692.2006.00090.x
9. Ye M, Shen J, Lin G, et al. Deep learning for person re-identification: A survey and outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2020; 44(6): 2872–2893. doi: 10.1109/TPAMI.2021.3054775
10. Manlik O, McDonald JA, Mann J, et al. The relative importance of reproduction and survival for the conservation of two dolphin populations. *Ecology and Evolution* 2016; 6(11): 3496–3512. doi: 10.1002/ece3.2130
11. Jefferson TA, Karczmarski L. *Sousa chinensis*. *Mammalian Species* 2001; (655): 1–9. doi: 10.2307/0.655.1
12. Liu L, Ouyang W, Wang X, et al. Deep learning for generic object detection: A survey. *International Journal of Computer Vision* 2020; 128(2): 261–318. doi: 10.1007/s11263-019-01247-4
13. Hinton GE, Vinyals O, Dean J. Distilling the knowledge in a neural network. *ArXiv E-print* 2015; arXiv:1503.02531. doi: 10.48550/arXiv.1503.02531
14. Cheng D, Gong Y, Zhou S, et al. Person re-identification by multi-channel parts-based CNN with improved triplet loss function. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 27–30 June 2016; Las Vagas, NV, USA. pp. 1335–1344. doi: 10.1109/CVPR.2016.149

15. Zheng L, Zhang H, Sun S, et al. Person re-identification in the wild. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 21–26 July 2017; Honolulu, HI, USA. pp. 3346–3355. doi: 10.1109/CVPR.2017.357
16. Luo H, Jiang W, Gu Y, et al. A strong baseline and batch normalization neck for deep person re-identification. *IEEE Transactions on Multimedia* 2020; 22(10): 2597–2609. doi: 10.1109/TMM.2019.2958756
17. Wang G, Yuan Y, Chen X. Learning discriminative features with multiple granularities for person re-identification. In: Proceedings of the 26th ACM International Conference on Multimedia; 22–26 October 2018; Seoul, Republic of Korea. pp. 274–282. doi: 10.1145/3240508.3240552
18. Sun Y, Zheng L, Yang Y, et al. Beyond part models: Person retrieval with refined part pooling (and A Strong Convolutional Baseline). In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y (editors). *Computer Vision—ECCV 2018. Lecture Notes in Computer Science*, Proceedings of 15th European Conference on Computer Vision (ECCV); 8–14 September 2018; Munich, Germany. pp. 501–518. doi: 10.1007/978-3-030-01225-0_30
19. Zhao H, Tian M, Sun S, et al. Spindle net: Person re-identification with human body region guided feature decomposition and fusion. In: 2017 Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 21–26 July 2017; Honolulu, HI, USA. pp. 907–915. doi: 10.1109/CVPR.2017.103
20. Liu C, Zhang R, Guo L. Part-pose guided amur tiger re-identification. In: Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW); 27–28 October 2019; Seoul, Korea (South). pp. 315–322. doi: 10.1109/ICCVW.2019.00042
21. Krschens M, Barz B, Denzler J. Towards automatic identification of elephants in the wild. *ArXiv E-print* 2018; arXiv:1812.04418. doi: 10.48550/arXiv.1812.04418
22. Bouma S, Pawley MDM, Hupman K, Gilman A. Individual common dolphin identification via metric embedding learning. In: Proceedings of the 2018 International Conference on Image and Vision Computing New Zealand (IVCNZ); 19–21 November 2018; Auckland, New Zealand. pp. 1–6. doi: 10.1109/IVCNZ.2018.8634778
23. Weideman HJ, Jablons ZM, Holmberg J, et al. Integral curvature representation and matching algorithms for identification of dolphins and whales. In: Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops (ICCVW); 22–29 October 2017; Venice, Italy. pp. 2831–2839. doi: 10.1109/ICCVW.2017.334
24. Hughes B, Burghardt T. Automated visual fin identification of individual great white sharks. *International Journal of Computer Vision* 2017; 122(3): 542–557. doi: 10.1007/s11263-016-0961-y
25. Feng J. *Photo Identification of Chinese White Dolphin Based on Deep Learning* [Master's thesis]. Shantou University; 2020.
26. Deng J, Guo J, Zhou Y, et al. RetinaFace: Single-stage dense face localisation in the wild. *ArXiv E-print* 2019; arXiv:1905.00641. doi: 10.48550/arXiv.1905.00641
27. Redmon J, Farhadi A. YOLOv3: An incremental improvement. *ArXiv E-print* 2018; arXiv:1804.02767. doi: 10.48550/arXiv.1804.02767
28. Hermans A, Beyer L, Leibe B. In defense of the triplet loss for Person re-identification. *ArXiv E-print* 2017; arXiv:1703.07737. doi: 10.48550/arXiv.1703.07737
29. Russakovsky O, Deng J, Su H. et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision* 2015; 115: 211–252. doi: 10.1007/s11263-015-0816-y
30. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 27–30 June 2016; Las Vegas, NV, USA. pp. 770–778. doi: 0.1109/CVPR.2016.90