Estuarine, Coastal and Shelf Science 269 (2022) 107815

Contents lists available at ScienceDirect

Estuarine, Coastal and Shelf Science

journal homepage: www.elsevier.com/locate/ecss

Coastal fisheries resource monitoring through A deep learning-based underwater video analysis

Dian Zhang [a,b], Noel E. O'Conner [a], Andre J. Simpson [c], Chunjie Cao [b], Suzanne Little [a,**], Bing Wu [c,d,*]

[a] Smart Sensing group, Hainan University, Haikou, China
[b] Insight Centre for Data Analytics, Dublin City University, Dublin, 9, Ireland
[c] Department of Physical & Environmental Sciences, University of Toronto Scarborough, 1265 Military Trail, Toronto, Ontario, M1C1A4, Canada
[d] Department of Chemistry, University of California, Berkeley, Berkeley, CA, 94720, USA

ARTICLE INFO

Keywords:
Ocean survey
Deep learning
Remote underwater video sensing
Mask region based convolutional neural network

ABSTRACT

Unlike land, the oceans, although covering more than 70% of the planet, are largely unexplored. Global fisheries resources are central to the sustainability and quality of life on earth but are under threat from climate change, ocean acidification and over consumption. One way to analyze these marine resource is through remote underwater surveying. However, the sheer volume of recorded data often make classification and analyses difficult, time consuming and resource intensive. Recent developments in machine learning (ML) have shown promising application in extracting high level context with near human performance on image classification tasks. The application of ML in remote underwater surveying can drastically reduce the processing time of these datasets. In order to train these deep neural networks used in ML, it is necessary to create a series of large-scale benchmark datasets to test any proposed algorithm for this kind of specific imaging classification. Currently, none of the publicly available datasets in the marine vision research domain have sufficiently large data volumes to reliably train a deep model. In this work, a publicly available large-scale benchmark underwater video dataset is created and used to retrain a state-of-the-art machine vision deep model (MaskRCNN). This model is in turn applied into detecting and classifying underwater marine lives through random under-sampling (RUS), and achieves a reasonably high average precision (0.628 mAP), indicating great applicability of this dataset in training instance segmentation deep neural network for detecting underwater marine species.

1. Introduction

Arguably, the most important ecosystem on earth, the ocean is central to multiple geophysical processes such as regulating global temperature, driving weather patterns, and sustaining a wealth of both living and nonliving resources. Due to the tremendous difficulties and costs of exploring the ocean, more than 80% of the oceanic area, according to the National Oceanic and Atmospheric Administration (NOAA), remains under-surveyed or unmonitored. (Kim and Mauborgne, 2005). Owing to the increasing pressure of climate change (Rahmstorf, 2002), overfishing (Jackson et al., 2001) as well as the overpopulation of human society (Samir and Lutz, 2017), effective management of the remaining oceanic resources (e.g. fishery, underwater mining) become quintessential to a sustainable future. In order to achieve this goal, it is necessary to firstly develop surveying capabilities to evaluate and monitor these local underwater resources. For example, real-time local monitoring could be provide a much needed tool on which to dynamic modify fishing quotas during a season.

Over recent years, underwater video monitoring systems, have gained significant attention and been deployed for coastal/marine ecological studies. With the rapid developments in technology such as high resolution digital cameras, high volume data storage and long range data transmission, the long term deployment of remote underwater video (RUV) sensing systems has been realized. Since the early introduction of RUV systems by H. Barnes around 1950 to survey marine activity along the Scottish coastline (Barnes, 1952), these underwater

* Corresponding author. Department of Physical & Environmental Sciences, University of Toronto Scarborough, 1265 Military Trail, Toronto, Ontario, M1C1A4, Canada.
* Corresponding author.
E-mail addresses: suzanne.little@dcu.ie (S. Little), friedrichbing.wu@utoronto.ca (B. Wu).

https://doi.org/10.1016/j.ecss.2022.107815
Received 11 July 2021; Received in revised form 5 March 2022; Accepted 7 March 2022
Available online 9 March 2022

monitoring networks have been frequently used in marine and coastal ecosystem studies (Condal et al., 2012; Fedra and Machan, 1979; Fisher et al., 2016; Jan et al., 2007; Miller et al., 2019). Among them, the SmartBay pilot infrastructure in Ireland is particular famous for its sub-sea cabled long time observatory (Fig. 1) equipped with multiple underwater sensors and probes (Gaughan et al., 2019).

The development in hardware has led to the exponential growth in the overall volume of recorded data, which subsequently swamps the capability of current data processing and analysis. This is particularly problematic in the field of marine ecology, where most research is based on data-derived statistical analyses. As a result of both spatial and temporal complexity of natural environment over various scales, a reliable ecological monitoring program would require widespread and continuous monitoring which generates an enormous amount of data to be evaluated (Kelling et al., 2009). Hence, it is necessary to utilize advanced data-processing technique to handle the sheer volume of ecological data produced.

With the introduction of deep learning algorithms, significant development has been made in the application of machine learning in scientific research. Several deep neural network models have achieved considerable improvement in various fields, such as computer vision (Krizhevsky et al., 2017) and language processing (Graves et al., 2013). As for marine ecology, Olsvik et al. proposed a Convolutional Neural Network (CNN) with Squeeze-and-Excitation architecture for classifying fish images, and achieved very high accuracy for fish classification (Olsvik et al., 2019). Mahmood et al. trained a VGGNet based deep model to detect the appearance of coral in images collected by an AUV near the Abrolhos Islands(Mahmood et al., 2016). Other than the significantly increased computational resources (e.g. the use of GPU), the key factor for these successful application of deep neural network lie in the publicly available, large-scale, benchmark datasets.(Cui et al., 2016).

SmartBay Facility carries out an all-day non-interruptive continuous underwater survey using a high resolution video camera (1280 × 720 pixels and 30 fps frame rate), while its anti-fouling and anti-reflection lens coating significantly reduces the occurrence of image deterioration, providing an ideal resource for oceanic ecology studies (Gaughan et al., 2019). In order to build a benchmark dataset from these high resolution videos, all the videos captured must be processed and annotated. It has been widely accepted in the research community that a supervised deep learning algorithm will generally achieve acceptable performance with 5000 labelled examples per category and will match or exceed human performance when trained with a dataset containing at least 10 million labelled examples.(Goodfellow et al., 2016). Unfortunately, only a few existing datasets are publicly available in the marine vision research domain, particularly, large-scale datasets that can be used to train a deep model are missing.

In this work, a realistic, large-scale, fine-grained, underwater benchmark video dataset utilizing the SmartBay underwater marine observatory is create. This dataset containing 13,946 frames with 118,830 fish objects were annotated from a subset of the videos captured by the SmartBay Facility. A deep neural network Mask R–CNN model (He et al., 2017), was re-trained using this dataset and the re-trained model achieved relatively high fish detection and segmentation accuracy, suggesting great potential for trained deep neural network models to extract high level information from raw underwater videos automatically. Artificial intelligent algorithms can assist marine biologists in better understanding our ocean environment at a much higher spatial and temperate scale than possible using manual methods. To support open research and continued development, the dataset and trained model is publicly available at https://github.com/DianZhang/missfish. We hope the dataset created and the results obtained in this work can inspire researchers in relevant research domains to propose and evaluate novel algorithms, to drive underwater marine research in a collaborative fashion.

2. Methods and experimental

2.1. Methods

Deep networks have been shown to be successful for computer vision tasks because they can extract appropriate features while jointly performing discrimination (Deng, 2014). In this study, a state-of-the-art object segmentation deep neural network, Mask R–CNN (Mask Region based Convolutional Neural Network), was implemented based on a publicly available git repository (He et al., 2017). Mask R–CNN is a conceptually simple, flexible and general framework for object instance segmentation. The network can efficiently detect objects in an image while simultaneously generating a high-quality segmentation mask for each instance. This network has won several computer vision challenges, including classification tasks in ImageNet Large Scale Visual Recognition Challenge (ILSVRC) and Microsoft Common Objects in Context (COCO) competitions. Extending its predecessor Faster R–CNN (Ren et al., 2017), Mask R–CNN added a branch for predicting segmentation masks on each region of interest, in parallel with the existing branch for object classification and bounding box regression. The architecture diagram of Mask R–CNN is shown in Fig. 2. It has been successfully used for human pose estimation (Guler et al., 2018), cell tracking (Tsai et al., 2019), face detection (Lin et al., 2020) amongst others. However, existing models were trained using the generic ImageNet and COCO datasets and, thus, it cannot be used directly to detect and segment underwater objects. Hence, a Mask R–CNN model is re-trained using the benchmark data created to evaluate the applicability of this dataset as well as the model itself.

To evaluate the performance, mean Average Precision (mAP) over all frames is used. It is widely used for evaluating and comparing object

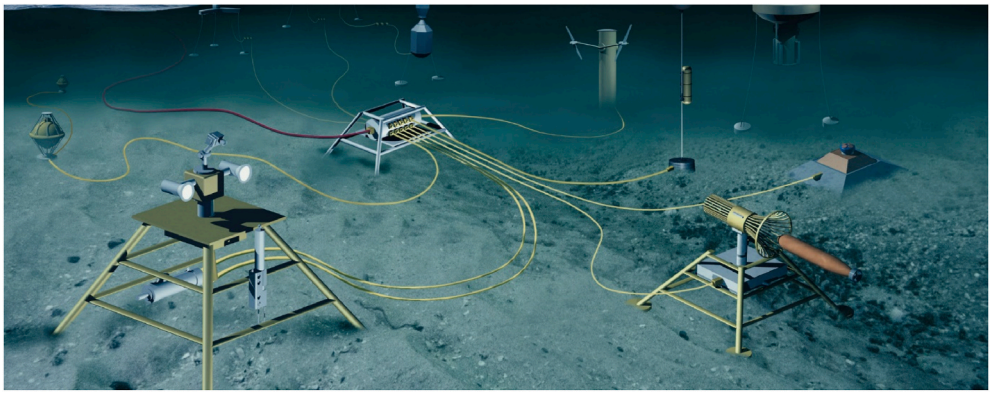

Fig. 1. Illustration of the SmartBay sub-sea observatory. A high definition Kongsberg pan-tilt-zoom camera system is installed (left bottom) and connected to the data and power hub (center).

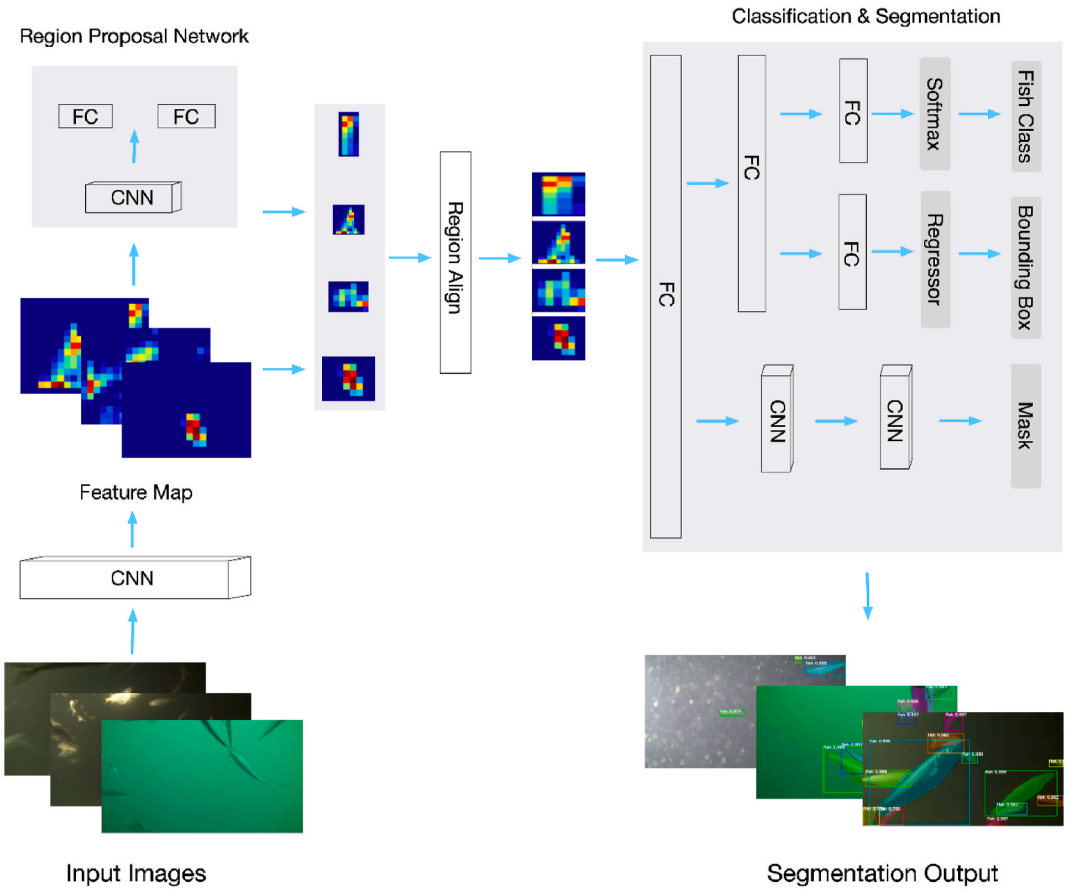

Fig. 2. The illustration of Mask R–CNN deep learning model.

detection and segmentation algorithms in computer vision communities. AP summarizes a precision-recall curve as the weighted mean of precisions achieved at each threshold. Here the threshold is defined as the intersection over the union (IoU), which measures the area of the model's output with the ground truth region divided by the total area of both. Following the common standard, the IoU threshold is set to 0.5 (AP@IoU=0.5) in this work. For Mask R–CNN and many other object localization methods, a proposed Region of Interest (RoI) is considered positive if it has IoU with a ground-truth region of at least 0.5. The mask loss, a binary cross-entropy loss that penalizes wrong per-pixel binary classifications, is defined only on positive RoIs. The mask target is the intersection between an RoI and its associated ground-truth mask. The range of average precision is between 0 and 1, where 1 means a perfect detection (alignment of the segmented regions).

2.2. *Experimental*

2.2.1. *Dataset creation*

Although over 400,000 video clips (each 2 min long) have been recorded at SmartBay facility so far, for the purpose of this study, a small portion of the available videos was selected by a half normal distribution method in a reverse chronological order to have a better representation of recent data. These videos were then manually filtered to exclude annulled videos (video without objects) that left only 45 videos in this dataset. The remaining videos were subsequently classified based on the capturing time of the video. In order to create a training dataset, every frames of the selected sample video were manually applied with two sample mask annotations (Fig. 3). It is worth noting that the sheer number of the moving subjects often complicates the annotations, while some variable environmental conditions like lighting conditions, also make this manual process challenging. Furthermore, the benchmark dataset was purposely constructed containing low, relatively low,

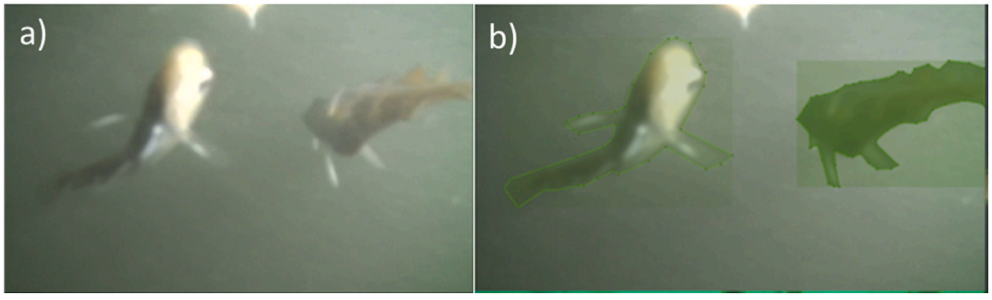

Fig. 3. The example of recorded video frame (a) and annotated video frame (b).

normal, and high visibility data. The authors have tried their best to build a benchmark data with as much variation as possible to reflect the true nature environment.

Furthermore, though various forms of marine life were present, the majority of these sea creatures captured in the these videos appeared to be fish. The occurrences of other types, such as crabs, are too sporadic to be used for any type of analysis. To create the ground truth, several annotation tools, including LabelMe, CVAT, LabelImg and VGG Image Annotator, were evaluated. However, the open source Microsoft VoTT were found to be most suitable for this task. The tool supports annotating both images and video frames with bounding box and polygon masks. For each video, three frames per second were annotated. This is due to the fact that fish movement can be very fast, especially when a fish is close to the camera lens. Three frames per second can provide fine-grained information for more sophisticated analysis, such as fish trajectory analysis.

As shown in Table 1, 13,946 frames from 45 videos and 118,830 objects (118,659 fish and 171 crab) were annotated in total. Due to the lack of specialist marine biology knowledge, labelling of fish species were not performed in this dataset, but could certainly be incorporated in future work. To comply with the machine learning data requirements, the annotated data is split into training, validation and test datasets. The first 30 videos recorded earlier in the chronological order were set as the training set, the following 10 videos were set as the validation set and the last 5 videos are used as test sets.

2.2.2. Deep neural network training

Initial experiments showed that existing pre-trained Mask R–CNN models do not perform well on underwater dataset. This is primarily because the datasets used to train existing Mask R–CNN models do not contain underwater images and underwater fauna. Thus, models need to be re-trained to detect fish. Utilizing existing models, the weights of the pre-trained model using the COCO dataset (Lin et al., 2014) is set as the initial network weights. Some parameters were adjusted to fit our dataset and hardware platform (see below). For comparison, a Mask R–CNN model was also trained from scratch. A script was implemented to convert VOTT annotation output to the input format that Mask R–CNN requires.

There are a number of hardware and data dependent parameters in Mask R–CNN that need to be set manually (detailed definitions of these parameters is described in previous work (He et al., 2017)). Since many of the fish in the dataset appear small, the *RPN_ANCHOR_SCALES* is set to (16, 32, 64, 128, 256). According to the frame size of the data, *IMAGE_MIN_DIM* and *IMAGE_MAX_DIM* are set to 512 (shorter edge is 512 pixels) and *TRAIN_ROIS_PER_IMAGE* is set to 64. The values of *IMAGES_PER_GPU*, *STEPS_PER_EPOCH*, *GPU_COUNT* and *VALIDATION_STEPS* are hardware dependent. Based on our experiment machine (Intel Core i7-4930K CPU with 32G DDR3 RAM) equipped with an NVidia TITAN X GPU card (3584 CUDA cores and 12G GPU memory), the following values are used $IMAGES\_PER\_GPU = 1$, $STEPS\_PER\_EPOCH = 200$, $GPU\_COUNT = 1$. Though some hyper-parameters in Mask R–CNN should also be tuned, it was found that most of the default values are robust enough for segmentation tasks. Several initial learning rates were chosen empirically but 10^{-5} was found to achieve the best performance. All the remaining parameters are set to default values.

Table 1
Statistics of the annotated dataset.

Total no. of videos	45
Total no. of frames	13946
Total no. of objects	118830
Fish instances	118659
Other instances (crab)	171
Ave. no. of instances per video	2637
Ave. no. of frames containing instances per video	310

3. Results

As described previously, in order to avoid overfitting, the dataset is split into training, validation and test sets. The corresponding number of frames and the number of fish in each set is listed in Table 2. A default resnet 101 model was used as the backbone of the network as the graphics card has sufficient RAM to support this model. For the purpose of comparison, two Mask R–CNN models were trained, one from scratch (model-scratch) using randomly assigned weights and the other was re-trained (model-retrain) using weights that were trained on MS COCO dataset. Both models were trained for 120 epochs with learning rate decay by a factor of 10 at the 50th and 80th epoch. Running on our machine, it takes approximately 50 min to finish an epoch. Initial experiments showed that any fine-tuning of the last few layers of model-retrain did not provide sufficient accuracy, thus, all layers of model-retrain model were systematically trained again with MS COO dataset.

The training and validation losses from both models have a similar trend. As shown in Fig. 4a, the training losses decreases rapidly at the beginning of the training until *c.* 40 epochs, from which the improvement slows down and the changing curve becomes smooth. In the case of validation dataset, the loss fluctuated throughout the whole training for both models as a result of the reduced frame number and object number. However, the overall decreasing trends of the loss in the training of validation dataset can be found in both models.

The accuracy curves of these models trained with validation and test datasets are shown in Fig. 4b. Though the performance of both models are improving over time, the accuracy of model retrain has a more rapid increase at the beginning of the training, and the improvements gradually fade away and finally keep at 0.6 mAP after 50 epochs. On the other hand, the improvement in the training accuracy of model scratch is significantly smaller regardless of the dataset used. It is worth noting that the loss and accuracy are not linearly correlated. The loss takes both the classification error and the disparity between the segmentation mask and the ground truth into account (per-pixel softmax), while the accuracy is calculated based on the target object detection rate only (with 50% overlap threshold in this work). Based on the above results, the model retrain achieved a far better performance than the model scratch. A more detailed comparison is shown in Table 3 (data in the table is calculated based on the results obtained from epoch 100 to 120). It can be found that the model retrain obtained much higher mAP scores (over 6.5 times) on both validation and test dataset.

Once a model is fully trained, it can be used for detecting fish in the surveyed raw video dataset. Two output sample videos are publicly available online. The trained models are publicly available as well, more info can be found at https://github.com/DianZhang/missfish. Fig. 5 shows a comparison between the number of detected fish by the model retrain and the ground truth (10 validation videos and 5 test videos).

4. Discussion

As mentioned before, preliminary experiments shows that pre-trained machine vision models, such as YOLOv2(Redmon and Farhadi, 2017), as well as Mask R–CNN, cannot be applied directly to underwater videos. This is because these models were trained with very different datasets (ImageNet, MS COCO or Open Images Dataset) that do not contain underwater images.

All these pre-trained models cannot detect or segment any fish instances in any of datasets in this study. Though FgSegNet, a foreground

Table 2
Details of the training, validation and test dataset.

	Number of Frames	Number of Objects	Ave. No. Obj. per Fra.
Training	8985	79896	8.89
Validation	3409	29525	8.66
Test	1552	9238	5.95

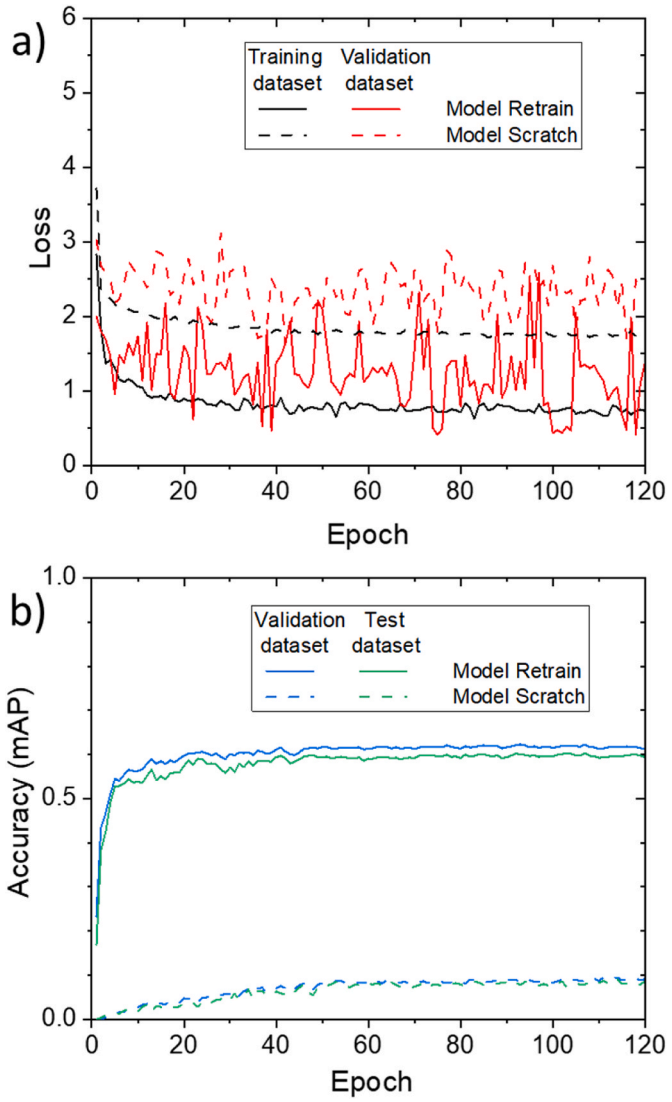

Fig. 4. The performance comparison between Model Retrain and Model Scratch in Loss (a) and Accuracy (b).

Table 3
Detailed comparison (between epoch 100 and 120) of training from model scratch and model retrained.

Epoch 100-120	Retrain	Scratch	Ratio
Mean Training Loss	0.73	1.75	2.41
Mean Validation Loss	1.03	2.28	2.22
Best Training Loss	0.61	1.72	2.80
Best Validation loss	0.41	1.63	3.99
Mean Validation mAP	0.617	0.090	6.88
Mean Test mAP	0.598	0.083	7.20
Best Validation mAP	0.623	0.095	6.56
Best Test mAP	0.603	0.092	6.58

segmentation model, did extract a few fish as foregrounds(Lim and Keles, 2018). However, it only achieved 0.0689 mAP for the validation datesets and 0.0419 mAP for the test dataset, which does not provide sufficient accuracy for underwater ecosystem analysis. Improvement has been observed from a re-trained FgSegNet model, however, improvements were very small (0.0719 validation mAP and 0.0637 test mAP).

Since the performed task is an instance segmentation from videos, the architecture of the network as well as the trained weights can be utilized. To retrain an existing model, a benchmark dataset needs to be

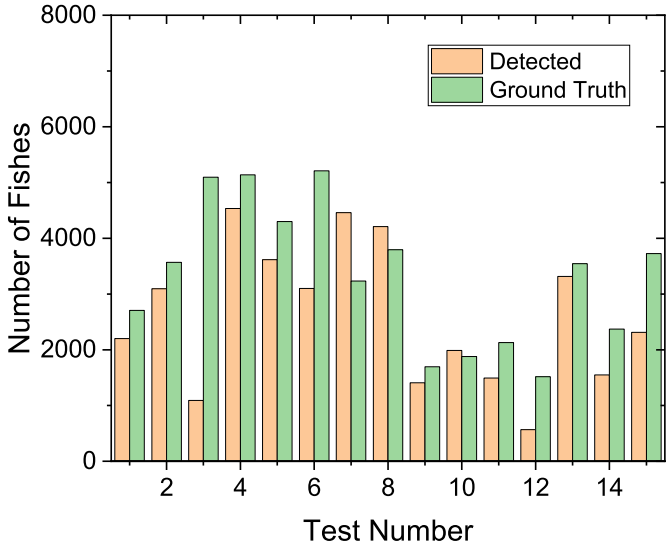

Fig. 5. The comparison between the number of fish detected by trained Mask R–CNN model and ground truth. (Test number 1–10 are validation datasets, 11–15 are test datasets).

first created based on the underwater videos captured by the SmartBay Observatory. Hence, the created dataset is the most refined (mask annotations compare to bounding boxes in other datasets) and one of the largest publicly available underwater computer vision training datasets. The subsequent evaluation of this dataset by the instance segmentation based Mask R–CNN model shows retrained model achieved a significantly higher detection accuracy compared to the model without retraining, suggesting a good applicability of this dataset in training deep neural network in classifying marine objects.

Furthermore, after visual inspection of all outputs, the majority of the errors were found to be either missed-labeling due to the severe object overlapping, or mis-labelled fish parts (Fig. 6). While the former is due to the limitation of the model, the latter type of error is caused by the annotation of partial fish in the benchmark dataset created, which is unavoidable for annotating fishes appearing at the edge of the camera. However, the classification scores of such partial sightings are typically low and may be filtered out in future work. Unlike other datasets (e.g. Fish4Knowledge (Fisher et al., 2016)) which only denotate fish when they fully appeared in the field of view, the benchmark dataset created in this work is trying to provide as much information as possible. Thus, all these partially appeared fish were annotated that results in this type of error. However, this dataset is designed for developing systems that can be deployed in real world scenarios in which many fish would appear partially at the edge of the camera view. Furthermore, it is also found that the trained model can effectively be used to estimate fish density levels. Compared to a human analyst, which may take days to analyze a short video, the trained model in this work can process the raw video data in minutes. This may provide a much higher temporal sampling information to assist marine biologists in better understanding the biodiversity of the environment.

5. Conclusions

In this study, a publicly available large-scale benchmark underwater video dataset was created, and used to retrain a state-of-the-art machine vision deep model (MaskRCNN). This model is in turn applied into detecting and classifying underwater marine lives through random under-sampling (RUS) and achieves a reasonably high average precision (0.628 mAP). Furthermore, results obtained in this work clearly shows that state-of-the-art computer vision deep neural network can be effectively applied to any underwater datasets through a retraining with the

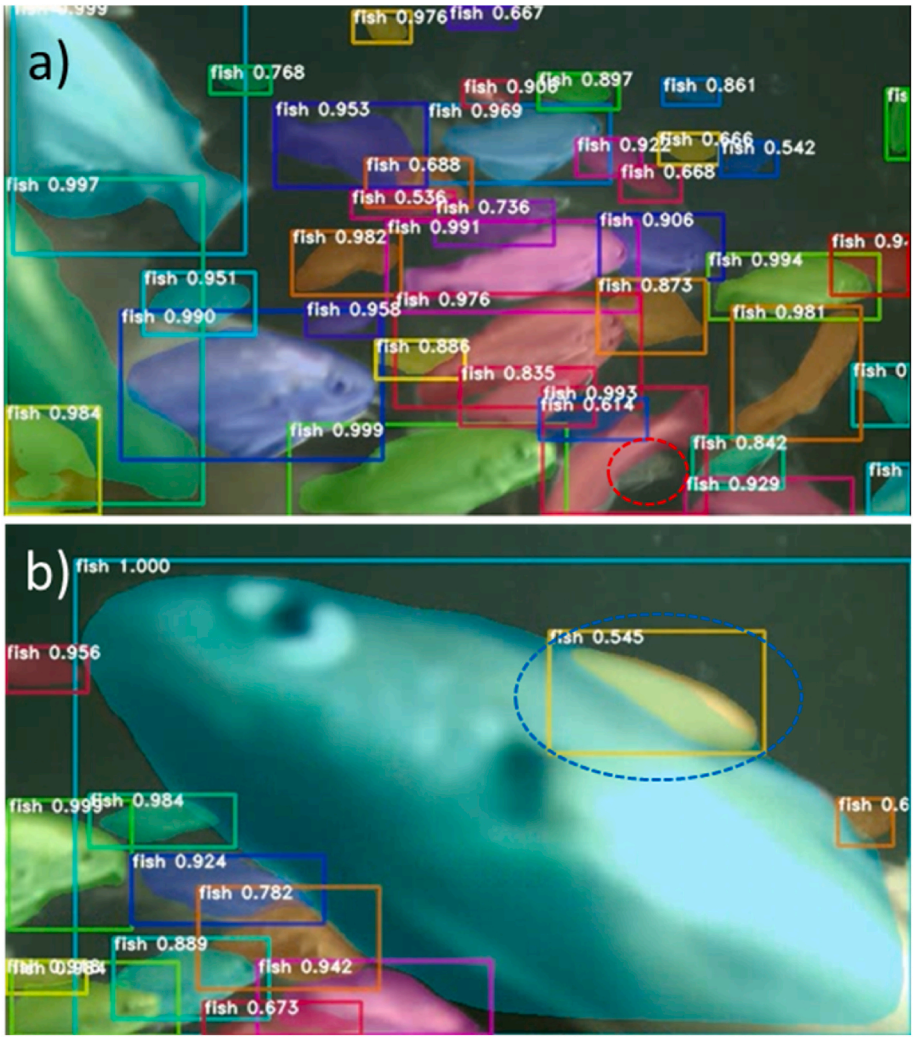

Fig. 6. Examples of some common errors made during the detection by the trained model. (a) strong overlapping of fishes (dashed red cycle); (b) mis-labelled fish part (dashed blue cycle). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

created benchmark dataset. In recent years, although machine learning techniques for computer vision processing have been improved significantly in both accuracy and diversity, the lack of a good benchmark underwater dataset hindered its application in marine ecological study. The creation of this public dataset in the underwater marine research domain would hopefully inspire more and more computer science researcher to develop new deep neural models to help marine scientists to better survey the marine resources. We hope with our initiation in creating a benchmark dataset, more similar datasets with higher accuracy can be created in the future, which in turn can drive computer science researchers around the world to develop, evaluate and test their new algorithms for studying marine environment.

CRediT authorship contribution statement

Dian Zhang: Conceptualization, Investigation, Methodology, Visualization, Writing – original draft. **Noel E. O'Conner:** Funding acquisition, Methodology. **Andre J. Simpson:** Writing – review & editing. **Chunjie Cao:** Funding acquisition. **Suzanne Little:** Conceptualization, Funding acquisition, Methodology, Resources. **Bing Wu:** Conceptualization, Methodology, Funding acquisition, Writing – review & editing, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research work is funded by Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289_P2, Natural Science Foundation of Hainan Province, fund number: 621MS017and the EU Horizon 2020 Marie Skłodowska-Curie Actions Cofund project (Grant No. 713279). It also received technical support from the National Infrastructure Access Programme, which is funded by the Marine Institute under the Marine Research Programme supported by Irish Government.

References

Barnes, H., 1952. Under-water television and marine biology. Nature 169, 477–479.
Condal, F., Aguzzi, J., Sarda, F., Nogueras, M., Cadena, J., Costa, C., Del Rio, J., Manuel, A., 2012. Seasonal rhythm in a Mediterranean coastal fish community as monitored by a cabled observatory. Mar. Biol. 159, 2809–2817.
Cui, H.G., Zhang, H., Ganger, G.R., Gibbons, P.B., Xing, E.P., 2016. GeePS: scalable deep learning on distributed GPUs with a GPU-specialized parameter server. In: Proceedings of the Eleventh European Conference on Computer Systems, (Eurosys 2016).

Deng, L., 2014. A tutorial survey of architectures, algorithms, and applications for deep learning. APSIPA Transactions on Signal and Information Processing 3, e2.
Fedra, K., Machan, R., 1979. A self-contained underwater time-lapse camera for in situ long-term observations. Mar. Biol. 55, 239–246.
Fisher, R.B., Shao, K.T., Chen-Burger, Y.H., 2016. Overview of the Fish4Knowledge project. Intel Syst Ref Libr 104, 1–17.
Gaughan, P., Berry, A., Malley, C.O., 2019. The dual roles of SmartBay, a multi-disciplinary subsea observatory delivering sustainable long term coastal marine observations and marine technology development. Ocean Ind.
Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. The MIT Press.
Graves, A., Mohamed, A.R., Hinton, G., 2013. Speech recognition with deep recurrent neural networks. Int Conf Acoust Spee 6645–6649.
Guler, R.A., Neverova, N., Kokkinos, L., 2018. DensePose: dense human pose estimation in the wild. Proc Cvpr Ieee 7297–7306.
He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask R-CNN, 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2980–2988.
Jackson, J.B.C., Kirby, M.X., Berger, W.H., Bjorndal, K.A., Botsford, L.W., Bourque, B.J., Bradbury, R.H., Cooke, R., Erlandson, J., Estes, J.A., Hughes, T.P., Kidwell, S., Lange, C.B., Lenihan, H.S., Pandolfi, J.M., Peterson, C.H., Steneck, R.S., Tegner, M.J., Warner, R.R., 2001. Historical overfishing and the recent collapse of coastal ecosystems. Science 293, 629–638.
Jan, R.Q., Shao, Y.T., Lin, F.P., Fan, T.Y., Tu, Y.Y., Tsai, H.S., Shao, K.T., 2007. An underwater camera system for real-time coral reef fish monitoring. Raffles Bull. Zool. 273–279.
Kelling, S., Hochachka, W.M., Fink, D., Riedewald, M., Caruana, R., Ballard, G., Hooker, G., 2009. Data-intensive science: a new paradigm for biodiversity studies. Bioscience 59, 613–620.
Kim, W.C., Mauborgne, R., 2005. Blue ocean strategy: from theory to practice. Calif. Manag. Rev. 47, 105-+.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. ImageNet classification with deep convolutional neural networks. Commun. ACM 60, 84–90.
Lim, L.A., Keles, H.Y., 2018. Foreground segmentation using convolutional neural networks for multiscale feature encoding. Pattern Recogn. Lett. 112, 256–262.
Lin, K.H., Zhao, H.M., Lv, J.J., Li, C.Y., Liu, X.Y., Chen, R.J., Zhao, R.Y., 2020. Face detection and segmentation based on improved mask R-CNN. Discrete Dynam Nat. Soc. 2020.
Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft COCO: common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.), Computer Vision – ECCV 2014. Springer International Publishing, Cham, pp. 740–755.
Mahmood, A., Bennamoun, M., An, S., Sohel, F., Boussaid, F., Hovey, R., Kendrick, G., Fisher, R.B., 2016. Automatic Annotation of Coral Reefs Using Deep Learning Oceans 2016 Mts/Ieee Monterey.
Miller, T.J., Hart, D.R., Hopkins, K., Vine, N.H., Taylor, R., York, A.D., Gallager, S.M., 2019. Estimation of the capture efficiency and abundance of Atlantic sea scallops (Placopecten magellanicus) from paired photographic-dredge tows using hierarchical models. Can. J. Fish. Aquat. Sci. 76, 847–855.
Olsvik, E., Trinh, C.M.D., Knausgard, K.M., Wiklund, A., Sordalen, T.K., Kleiven, A.R., Jiao, L., Goodwin, M., 2019. Biometric fish classification of temperate species using convolutional neural network with squeeze-and-excitation. Lect Notes Artif Int 11606, 89–101.
Rahmstorf, S., 2002. Ocean circulation and climate during the past 120,000 years. Nature 419, 207–214.
Redmon, J., Farhadi, A., 2017. YOLO9000: Better, Faster, Stronger, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6517–6525.
Ren, S.Q., He, K.M., Girshick, R., Sun, J., 2017. Faster R-CNN: towards real-time object detection with region proposal networks. Ieee T Pattern Anal 39, 1137–1149.
Samir, K.C., Lutz, W., 2017. The human core of the shared socioeconomic pathways: population scenarios by age, sex and level of education for all countries to 2100. Global Environ. Change 42, 181–192.
Tsai, H.F., Gajda, J., Sloan, T.F.W., Rares, A., Shen, A.Q., 2019. Usiigaci: instance-aware cell tracking in stain-free phase contrast microscopy enabled by machine learning. Software 9, 230–237.