DeepRecog: Threefold underwater image deblurring and object recognition framework for AUV vision systems

M. V. Pranav[1] · A. V. Shreyas Madhav[1] · Janaki Meena[1]

Abstract

Underwater explorations and probes have now become frequent for marine discovery and endangered resources protection. The decrease in natural light with an increase in water depth and the characteristic of the medium to absorb and scatter light pose crucial challenges to underwater vision systems. Autonomous Underwater Vehicles (AUVs) depend upon their imaging systems for navigation and environmental resource exploration. This paper proposes DeepRecog—an integrated underwater image deblurring and object recognition framework for AUV vision systems. The principle behind the image deblurring module involves a threefold approach consisting of CNNs, adaptive and transformative filters. The ensemble object detection and recognition module identifies marine life and other frequently existent underwater assets from AUV images and achieves mean Average Precision (mAP) of 0.95 and was found to be 6.42% more precise than YOLOv3, 8.43% more than FasterRCNN + VGG16 and 15.78% more than FasterRCNN. This framework was created with the purpose of providing real-time detection and recognition with minimal delay. The system can also be employed for former images acquired from AUVs and hopes to facilitate efficient solutions for marine image post-processing.

Keywords Underwater vision · Deep learning · Object recognition · Image deblurring · Autonomous underwater vehicles

1 Introduction

The development, exploration and protection of marine life and underwater resources has gained significant attention across the world due to the increase in climate changes and global warming. Recent implementations in the domain of marine research have resulted in the fruition of advanced autonomous and manually operated vehicles for transporting visual equipment for detection and recognition of necessary targets in underwater conditions. The field of underwater exploration is a state of constant development and innovation due to the inherent need of imaging processing and computer vision techniques for understanding visual information that are corrupted by a wide number of factors. Light attenuation, scattering, non-uniform lighting, shadows, colour shading, suspended particles, obscured vision due to the existence of

marine life are major contributors to this decrease in their ability to interpret valuable information from collected data.

1.1 Autonomous underwater vehicles

Autonomous under-water vehicles (AUVs) function completely on their own without the need of manual intervention and it is essential for them to have a viable perception of the elements in their surroundings. AUVs and their ability to extract valuable inferences from captured images is limited by the aforementioned factors that are characteristic to the medium. Figure 1 showcases the general operational flow of an AUV robot.

AUVs have now become increasingly relevant and serve as optimal options for water body research explorations. The range of depth, from shallower waters to extremely deep trenches, that can be serviced by AUVs make it appropriate for ocean research. The ability of AUVs to stay underwater for large and continuous periods of time makes them far more efficient than human divers. The sensory components and perception methods employed can be altered according to the task at hand without the requirement for expensive changes to the overall design of a AUV. Cameras

✉ A. V. Shreyas Madhav
shreyasmadhav.av2018@vitstudent.ac.in

[1] School of Computer Science and Engineering, Vellore Institute of Technology, Chennai 600127, Tamilnadu, India

Fig. 1 General process of AUV Operations
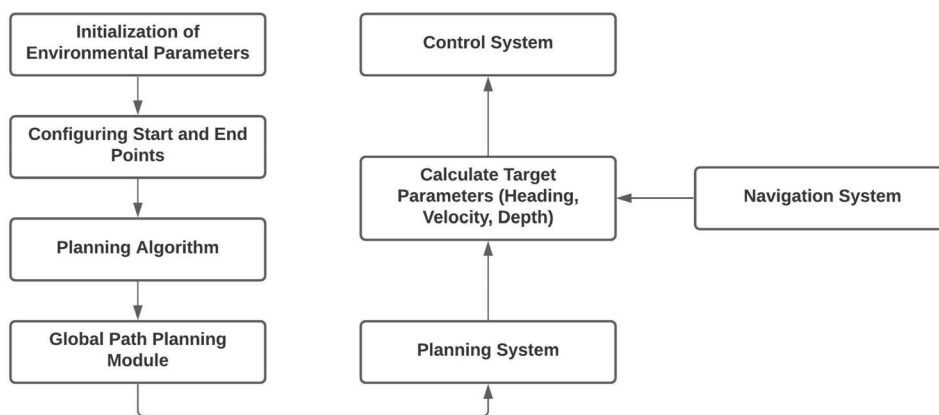
are employed in AUVs which explore regions of adequate illumination. Recent camera-based explorations for deeper ocean regions have been facilitated through strategically placed light strobes. The images may be affected by blur due to underwater disturbances and illumination deficiency, which has given rise to the field of image deblurration for underwater images.

1.2 Underwater image deblurring

Deblurring algorithms and methodologies form an integral part of the AUV vision system for the enhancement of captured images since the obtained data requires feasible pre-processing techniques in the majority of the cases. While light flashes from the visual equipment can enhance visibility, enhancement algorithms are still essential for underwater dark environments to enable autonomous object recognition. The retransformation of the obtained image to its original form remains a challenge due to the requirement to accommodate the blaring and distortion inflicted on the image. Deblurring processes are focussed on removing external and machine-based noise by estimating the blur kernel information and de-convoluting the image to obtain the ground truth representation. Both data-driven and traditional techniques have been employed in the past for this purpose. Recently, blind image deblurring techniques have gained traction in the field of image processing research due to their ability to restore the initial image with very little information on the attributes of the blur kernel. The reduction in peak signal-to-noise ratio (PSNR) within a considerable execution time is imperative for the real-time functioning of the system.

1.3 Underwater object recognition

Recent innovations in machine learning have propelled object localization and object recognition techniques to new heights. Object detection can be defined as the identification of the locations of required targets in an image and their dimensions. The recognition component depends heavily on classification modelling to categorise the target. Image classification models understand the visual features inherent to an image and assign a class label relevant to it. Object detection and recognition methodologies follow a twofold approach of isolating the region of interest from an image with diverse elements and classification of the recognised region to its appropriate label. The high amount of information and computational nature of image matrices has led to the development of deep-learning models that deploy extensive learning parameters and complex nodal architectures to traverse and understand the underlying information present in an image. Underwater object detection is dependent upon the quality of the camera and is usually combined with an image processing module to enhance the image quality. This paper proposes a novel framework for deblurring and object detecting underwater images obtained by AUVs to integrate the overall processes required for an AUV vision system.

2 Related works

The modern era of underwater imaging began with the development of electrical vision systems. The implementations of SONAR or camera-based imaging in AUVs and marine-exploration probes have been extant for quite some time and the need for employment of visual processing and recognition techniques are ubiquitous in marine research due to the unclear and noisy state of the medium. Applications requiring manual intervention have now become obsolete and automated recognition systems are taking over.

Techniques for reversal of distortions and degradations produced in the image due to the light-diminishing and scattering properties of the water bodies have led to innovative proposals which deal with contrast-stretching and adaptive thresholding based upon existing edge-detection operators like Sobel, Canny, Prewitt, etc. While limited range detection is still viable, the expansion of the visual recognition range can be expanded significantly with the introduction of appropriate deblurring strategies. The main focus of these

processes is to derive a proper visual representation by reducing the PSNR and SSIM [1]. Weighted guided filtering for deblurring to lessen halo artifacts can be propelled to the next level using gradient domain guided image filters that are focussed on blare restraining and boundary conservation [2]. Single image super-resolution of underwater images has also been proposed in the past using a set of low resolution and high-resolution compact cluster dictionaries. The removal of unwanted signals in the image, especially caused due to the suspended particles in the ocean water, was implemented using object detection and removal [3]. The two-fold approach was significant in removing the marine particles while preserving target object edges [4]. One significant breakthrough in removing the undesirable characteristics of colour distortions and visible noise was attained by the simplification of the Jaffe-McGlamery optimization algorithm by G. Huo et.al. Their approach was based on the derivation of a red-arc channel [5] prior to the estimation and transmission of background light. A simple and efficient low-pass deblurring filter was also proposed and the experimental results conclusively proved that their proposed algorithm was feasible for eliminating the influence of absorption and scattering [6]. Underwater image segmentation establishes itself as a reliable and stable preprocessing method [7] for enhancing the accuracy of target tracking and recognition. Segmentation algorithms in this field of research aim to solve the contour-deformation and edge-expansion problems in traditional methodologies. The modern segmentation algorithms are geared towards removing haze and improving object visibility [8].

Artificial object-based mean-shift tracking and template matching designs for underwater robots have been proposed with the combination of a novel weighted correlation coefficient employing colour and feature-based techniques to test the performance under various lighting conditions [9]. Their system was tested using an underwater robot platform yShark made by KORDI. Frameworks for AUV motion planning take into consideration both the self-dynamics of its actuators along the water-flow motion features [10]. The generation of vertices leads to an extension of controller action considered in previously existing literature where the circulation and location are considered as discrete values in time with optimum constraints achieved through multiprocessing. The increasing need for real-time data processing for onboard mission planning and adaptations in AUV route decisions due to the wide bandwidth requirements and data-intensive computations was the main purpose for the development of anomaly detection frameworks in the past [11]. The need for instant mitigation and response is crucial in dealing with situations that may cause damages or disastrous outcomes to the AUVs. The existing frameworks demonstrate their capability to side-scan SONAR datasets collected by AUVs where the identification of salient regions is performed by newly developed algorithms that are analogous to key-point matching and detection techniques in the field of image processing [12]. The framework also allows the transfer of obtained imagery for analysis by the operators and their relevant feedback. One prime example of an efficient qualitative navigation system was established by Memorial University [13]. Memory Explorer AUV enables path following and localization where a globally referenced position estimate is not necessary for its operation along the trained route.

Several object-detection algorithms are currently being applied for ocean exploration, employing contour segmentation and border-mapping techniques to locate objects and realise the target position [14]. Object detection data models and datasets are a crucial requirement in the field of underwater resource tracking and navigation. UDD is one such underwater open-sea farm object detection dataset that consists of images classified into three labels—scallop, sea cucumber, and sea urchin. It is one of the first datasets collected in a real open-sea farm with close to 2227 images. The paper also proposed a novel GAN algorithm (Poisson GAN) to combat class-imbalance issues in UDD [15]. Other object detection algorithms are usually built upon Convolutional Neural Networks (CNN). Deformable CNNs [16] pre-process underwater images to increase contrast and remove deviation of colour. ResNet-101 was utilised as a sub-network for feature extraction using deformable convolutional models and showcased prominent feature extraction improvements. Video-based object detection and summarization techniques have also established themselves as significant contributors to the design of technologically well-equipped underwater vision systems [17–22]. With the development of both image enhancement and object detection networks for marine resource recognition, it is imperative to understand the correlation between these models. Changes in the parameters defining image quality after enhancement processes and the accuracy achieved in object detection were carried out. An increase in accuracy on the image-enhanced dataset was recorded but no direct statistical correlations were established between the parameter changes and final detection accuracy [23]. ResFeats [24] based feature extraction processes have also proven much more efficient compared to CNNs for underwater image classification tasks. The absence of an integrated solution to perform deblurring and object recognition on the same platform has been a path of extensive research and this paper proposes a complete framework for realtime and post processing of image data in AUV vision systems.

Based upon careful consideration of existing works, an integrated system consisting of functional deblurring and object detection, specific to underwater exploration, was found to be lacking. The significant contributions of this work includes a triadic deblurring approach coupled with an

ensemble object detection module. This approach provides the combinational benefits of close to real-time results and visual robustness of clarifying images of disadvantageous resolutions.

3 Proposed system

The DeepRecog framework follows a hybrid approach of combining image dehazing and underwater object recognition for enhancing AUV image interpretations. The novel framework and its process flow are depicted in Fig. 2. The functioning of the framework is set in motion once the image is captured by the AUV vision system. The attained visual data is passed through a custom layered deep-learning model for deblurring and the processed image is made feasible for object detection. An ensemble object detection module has been built to predict target boundaries and their classes by obtaining a weighted average of two pre-trained models subjected to transfer learning (YOLOv5 and MR-CNN). As the system executes a dual model approach for underwater object detection and deep-learning-based image processing, the final recognition outputs of our DeepRecog framework provide a concise and visually accurate solution, disregarding irrelevant objects. This will alleviate the vulnerabilities and weak points of existing AUV vision recognition frameworks.

3.1 Deblurring module

The deblurring algorithm follows a triadic approach—an end to end transmission map is estimated using CNNs, colour deviation is removed based upon white balance parameters, and the final image is de-noised using hybrid wavelets and directional filter banks. The CNN is focussed on feature extraction, non-linear regression, local extremum and multi-scale mapping. The feature extraction is carried out by three kernels of different sizes to extract multi-scale

features. The final output is compressed by the Maxout activation function and is normalised using a bilateral rectified linear unit (BReLU). Unrealistic colour deviations can be amended based upon light estimation and colour correction. The implementation of the initial CNN for single image deblurring for the underwater images follows the principle of DehazeNet [25]. The first step is to calculate the lighting of the image for every colour channel with the use of Minkowski p norm. The unavailability of red components and properties of white objects underwater are some factors considered during the selection of the p value [26]. For colour corrections, we use comprehensive comparison for severely colour deviated underwater images [27]. The colour deviation is corrected iteratively by finding grey pixels and comparing their deviations. The colour corrected and blur free images are combined together Laplacian pyramids are utilised to obtain an amalgam between the colour corrected and blur free image. Each input image is modelled into different scales and every normalised weight map is calculated. The final image before edge detection is obtained by:

$$R^I(x, y) = G^I\{W(x, y)_k\} L^I\{I(x, y)_k\},$$

where I shows the pyramid level count, W is the normalised weight map, G{W} is its Gaussian version and L{I} is the Laplacian form of I The edge detection component of the module comes into play in the form of subjecting the image to HWD Transformation [28]. The HWD transformation disintegrates the images into L levels and the high frequency sub banks are subjected to directional filter banks. Texture and contour features are more accurately captured by the HWD transformation.

3.2 Ensemble detection module

Elementary object-detection algorithms were not as systematic as we want them to be today. To detect an object,
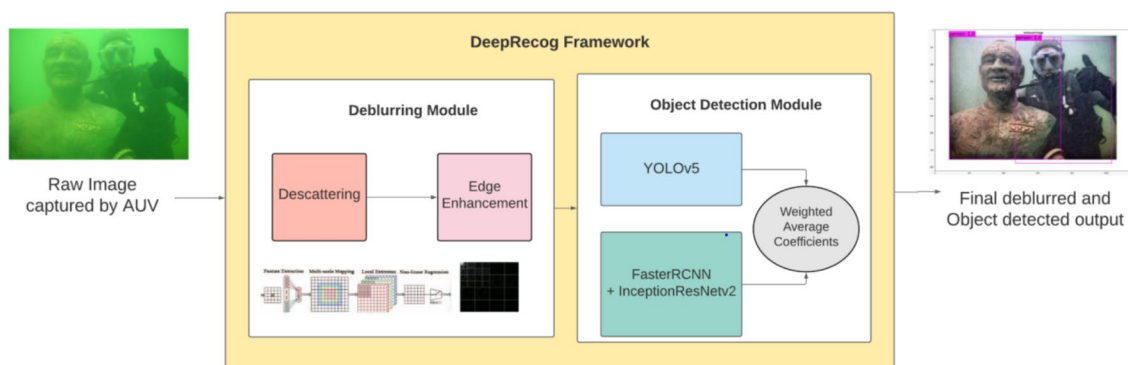

Fig. 2 DeepRecog architecture

the methodology involved implementing a classifier for the particular object, and estimate its closeness at several locations of the image. Many of the said algorithms used a sliding-window style to run the classifier at uniformly spaced regions over the entire image matrix. More recent trends include the use of R-CNNs that employ the use of region proposal methods to initially generate probable bounding boxes. The said classifier was limited to running over these boxes for recognition, rather than the entire image. Post-processing techniques for filtering and increasing the accuracy of the boxes, as well as the removal of duplicate boxes were included.

Most of the popular object-detection algorithms have one main drawback—speed for real-time object detection. YOLO [29] re-defines object detection as an uncomplicated regression model. The naming 'You Only Look Once' is administered literally, as the system only looks once at the image to predict the objects. The consolidated model has multiple advantages over earlier methods and is specifically optimised for detection performance. The decreased processing time can be attributed to the fact that object detection is defined as a regression problem, which negates the need for a complex pipeline. In this paper, we implement a weighted ensemble object detection module implementing two recently established object detection models (YOLOv5 and a hybrid FasterRCNN + Inception-ResNet V2). The weighted ensemble structure allows us to combine different structural models into the same module. The final prediction region is obtained from the models by structuring them as coefficient weighted ensembles trained independently.

3.2.1 YOLOv5

YOLO aims at the image globally, rather than region-restricted techniques mentioned earlier. The entire image is understood during training and testing to encode the correspondent data of the objects, as well as their other visual attributes. It generates a generic rendition of objects and their boxes. This step also involves the usage of non-maximal suppression and Intersection-over-union to excise duplicate boxes. YOLOv5 set the benchmark for object detection models very high. 4 models of YOLOv5 are publicly available, each having its own pre-trained weights on the COCO dataset. The said dataset is not inclusive of objects/animals found underwater, which necessitates the need for transfer learning. Images of underwater marine life were taken from a variety of publicly available datasets. Training of YOLO models requires the annotations of each image that is the coordinates of the rectangle that encompasses the required object in the said image. While some datasets came along with their annotations, others required manual annotation via software like LabelImg that allows the user to manually select the coordinates of the object.

3.2.2 FasterRCNN + InceptionResNet V2

InceptionResNet V2 is a pre-trained convolutional network that has a depth of 164 layers and the ability to classify 1000 object categories robustly without the need for custom learning. The network, originally trained on a wide range of images from Imagenet (close to 1 million images), has attained rich feature characteristics and identification techniques. The image input size for the network is set at 299*299. Inception V2 [30] has gained attention due to its ability to widen the
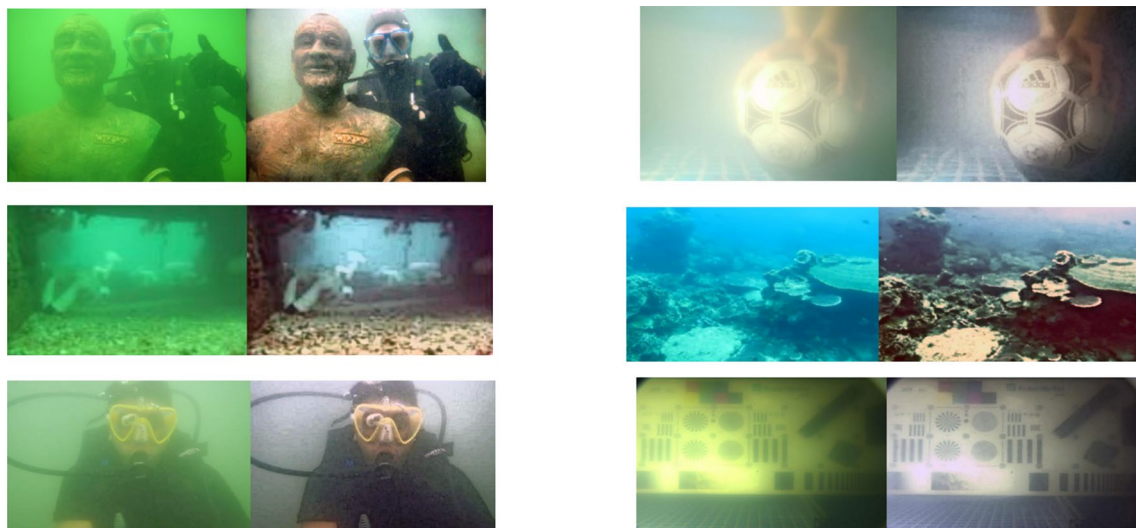

Fig. 3 Comparison of raw images with the DeepRecog deblurred images

architecture of the network rather than deepening it. The inclusion of residual nature into the original Inception module has proved beneficial in several past works. Residual Inception architectures outperform all similar Inception Networks that are implemented without residual connections.

Faster RCNN, the successor of Fast RCNN and the original RCNN, is one of the most renowned deep convolutional networks with an Object Detection component and a Region Estimation Network (REN). The region proposals are predicted by a separate network instead of the implementation of a selective search algorithm on the feature maps. The faster neural network in place of the original algorithm is one of the main improvements of the Faster RCNN in comparison to the previous such object detection algorithms. The Region Estimation/Proposal Network is another major addition contributed by the Faster RCNN development. The feature maps are scaled down to decreased dimensions by a sliding window in the final stages of the initial CNN. Multiple likely regions are generated at each location of the sliding window based on default bounding boxes. Different sized boxes are tested for their probability of encompassing an object and their coordinates. Softmax probability is considered for the conclusion of the best bounding box most likely to contain the object. The Region proposal network works primarily towards estimating the box coordinates and does not classify the bounded box objects. If a certain threshold of probability is passed by the bounding box it is proposed as a region of interest. Once the region of interest has been finalised, they are fed into the main network of pooling and fully connected layers of the Fast RCNN. The final layer is a Softmax classification and a bounding box regressor is also in place. Tensorflow's implementation of the Fast RCNN model with Inception ResNet is one of
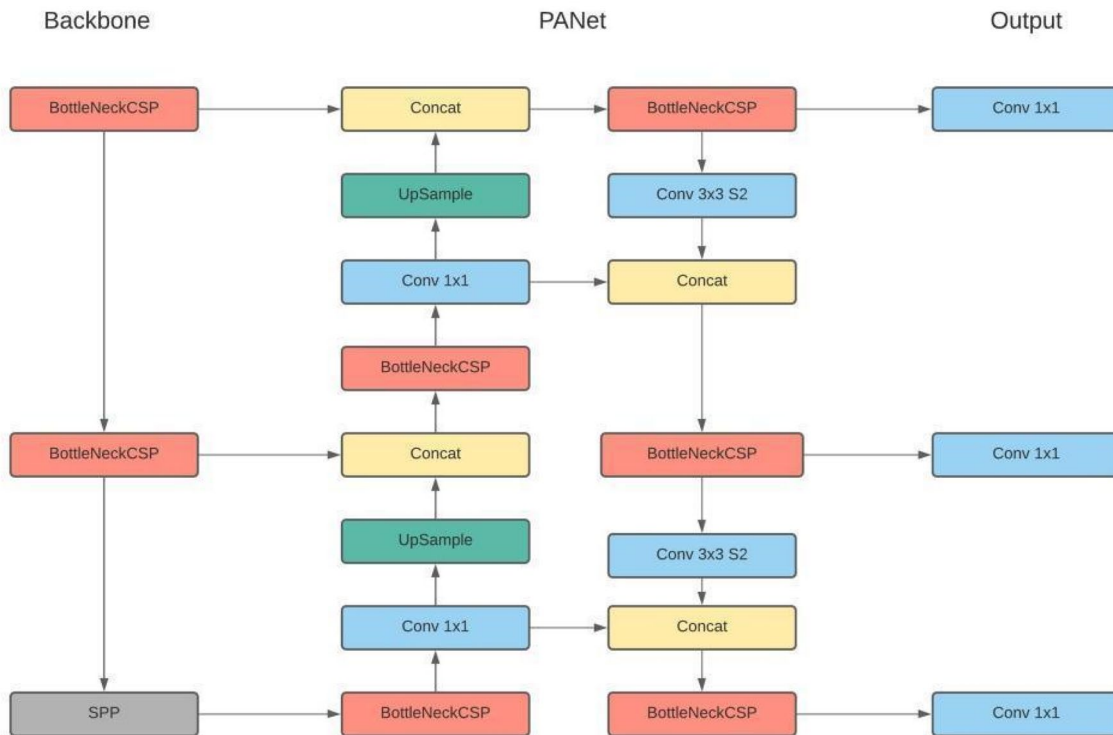

Fig. 4 YOLOv5 overview

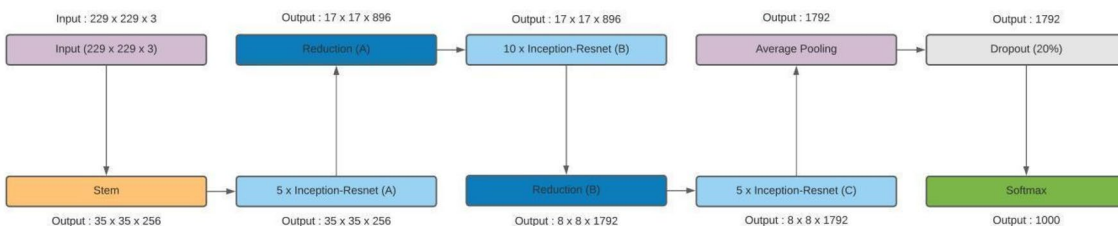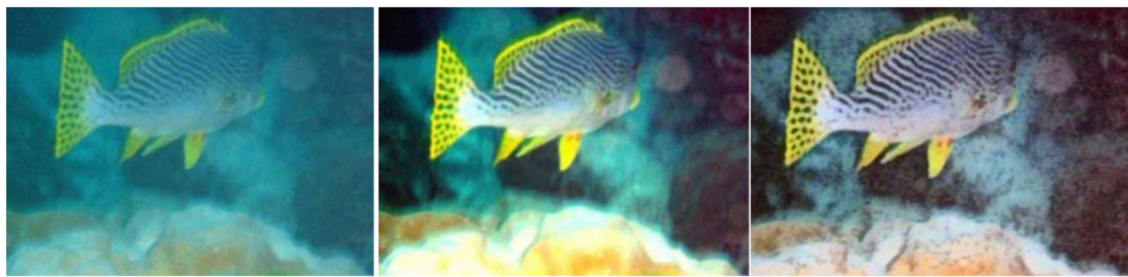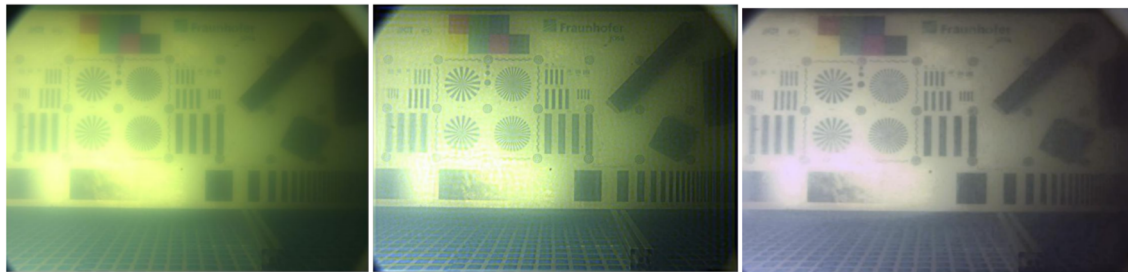**Fig. 5** InceptionResNetv2 overview

(a) Captured Image   (b) Deblurred Image of Iqbal et.al [20]   (c) DeepRecog deblurred image

(b) Captured Image   (b) Deblurred Image of Stephen et.al [21]   (c) DeepRecog deblurred image

(a) Captured Image   (b) Deblurred Image of Y-T.Peng [22]   (c) DeepRecog deblurred image

(a) Captured Image   (b) Deblurred Image of Hitam et.al [23]   (c) DeepRecog deblurred image

Fig. 6 DeepRecog deblurring effects comparison with existing works

Fig. 7 DeepRecog object detection metrics

Table 1 Object detection model comparisons

Model	Mean average precision (map)	Difference in mAP (%) over extant models
DeepRecog Ensemble	0.95	–
YOLOv3	0.89	6.42
Modified YOLOv3	0.87	8.43
FasterRCNN + VGG16	0.87	8.43
Faster RCNN with region network proposal acceleration	0.83	12.63
RCNN with bounding box regression	0.81	14.73
Faster RCNN	0.80	15.78
Fast RCNN with single value decomposition	0.79	16.84
Fast RCNN	0.73	23.15
Deformable Parts Model (Baseline)	0.67	29.47

their most accurate models and hence. Has been considered as part of our ensemble structure. Once the classification has been made by the model, the object is bound in the image along with its appropriate classified label.

4 Experimental results

4.1 Dataset

The dataset comprises 7000 jpeg images of 7 different underwater specific categories namely—humans, fishes, jellyfish, starfish, sharks, tortoises and coral-reef. The images were compiled from various open-source image datasets with Open Images Dataset v6 (Google) as the primary contributor. A twofold approach was employed to improve the robustness of the classification model. The first approach focussed upon obtaining significant naturally blurred images as well as synthetic blurred images with randomised noise functions. The second approach was obtaining non-blurred underwater images belonging to the relevant classes. For the training process, the blurred images were processed using our aforementioned deblurring module before feeding it into the network while the non-blurred images were used

directly. A majority of the images were obtained along with their annotations whereas the remaining dataset was manually annotated using LabelImg (Figs. 3, 4, 5).

4.2 Implementation

The entire framework was coded on a Ryzen 5 3600 @ 3.6 GHz, 16 GB RAM PC. It is built upon MATLAB (deblurring module) and Python (object detection) and integrated using the Matlab Engine API. To evaluate the higher visual enhancement of our blur-removal algorithm, an extensive comparison was drawn among recent works in underwater image deblurring. Figure 6 showcases the visual enhancements of our model in comparison with the ground truth and other existent works. Hence, it can be inferred that our deblurring module provides a much more visually refined output suitable for further image operations (in this case, object detection).

Since the availability of a CUDA compatible GPU is highly beneficial, a Google Colab environment was used for training. The general method to calculate the value of Average Precision (AP) is to estimate the area under the Precision-Recall curve. mAP can be determined as the average of AP. Talking in particular to object detection, the mAP
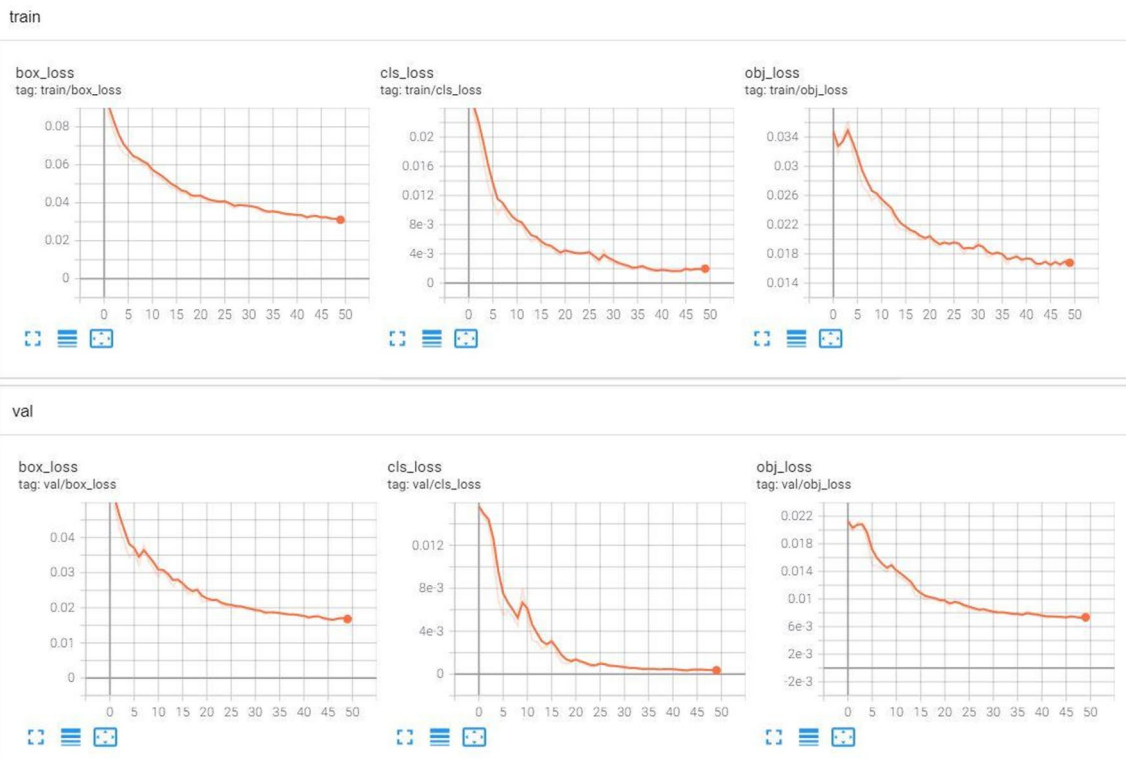
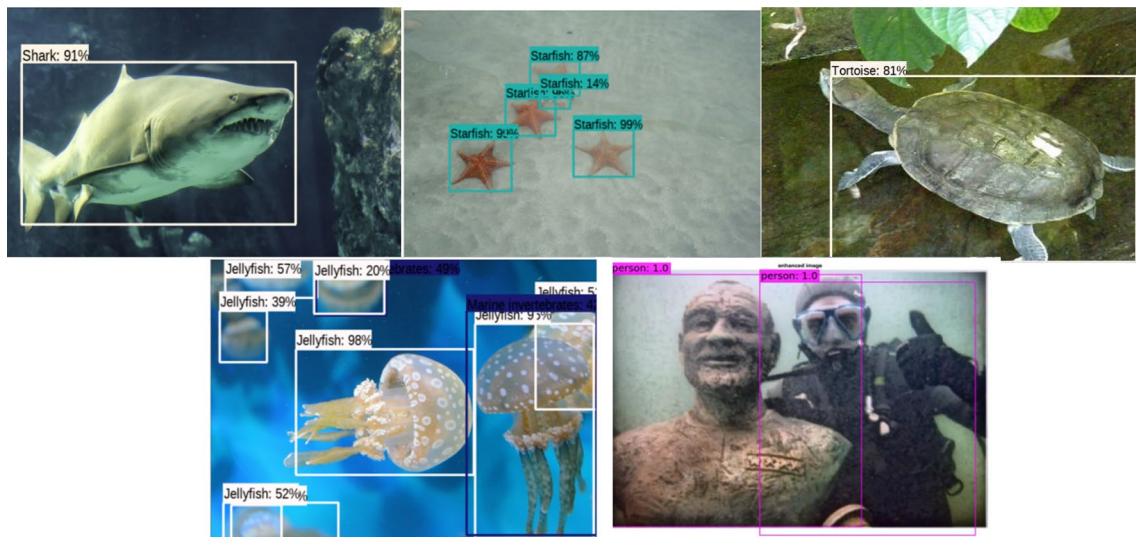**Fig. 8** Training and validation curves of the combinational detection module

Fig. 9 DeepRecog object detection metrics

score is calculated by computing the mean AP overall IoU thresholds, depending upon the specific parameters of the model. A total of 1000 images (per class) were trained on the model for 50 epochs. Figure 7 shows the metrics which threw a mAP accuracy score of 0.95, precision of 0.88 and recall of 0.0.93. Table 1 draws a comparative analysis of

our ensemble with existing models [31–33] for underwater object detection.

The training and validation losses of the model can be seen in Fig. 8. In both cases, it can be observed that loss is almost negligent as both graphs tend towards zero. Figure 9

showcases the final results of our DeepRecog framework that recognises the target objects present in the image.

Based upon analysis of the training and validation curves of the combinational model, we can conclude that the model is a good fit to the dataset without any sign of overfitting or underfitting. The mAP (mean Average Precision) obtained by the object detection network of our DeepRecog Framework is 6.42% better than the closest state of the art model(YOLOv3) and is 29.47% better than the baseline model(Deformable parts model. Overall, the DeepRecog framework acts as an optimal addition to existing AUV vision by collaborating the processes of deblurration and underwater specific object detection.

5 Conclusion

DeepRecog accomplishes the combinational proposal of integrating deblurring and object detection into a single application entity focussed towards marine resource research and improving AUV vision. The deblurration system provided a water specific methodology for the removal of haze and noise while preserving the visual integrity of the original image. The novel object detection module for underwater items was 6.42% more precise than YOLOv3, 8.43% more than FasterRCNN + VGG16 and 15.78% more than FasterRCNN. The future scope of research may be directed towards accommodating illumination enhancement modules for deep-sea AUV vision systems. With recent advances in autonomous underwater transportation, the possible depth and range of underwater exploration have increased and the object detection system can be modified to include more categorical labels.

Declarations

References

1. D. Kim, D. Lee, H. Myung and H. Choi (2012) Object detection and tracking for autonomous underwater robots using weighted template matching, 2012 Oceans—Yeosu, Yeosu, Korea (South), pp. 1-5, doi: https://doi.org/10.1109/OCEANS-Yeosu.2012.6263501

2. Cheng, C., Sung, C., Chang, H.: Underwater image restoration by red-dark channel prior and point spread function deconvolution, 2015 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), Kuala Lumpur, Malaysia, 2015, pp. 110–115, https://doi.org/10.1109/ICSIPA.2015.7412173.

3. Iqbal, K., Odetayo, M., James, A., Salam, R.A., Talib, A.Z.H.: Enhancing the low quality images using unsupervised colour correction method. In: IEEE International Conference on Systems Man and Cybernetics (SMC). 2010, pp. 1703–1709.

4. Farhadifard, F. (2017). Underwater image restoration: super-resolution and deblurring via sparse representation and denoising by means of marine snow removal.

5. Peng, Y-T., Cosman, P. C.: Underwater image restoration based on image blurriness and light absorption. In: IEEE Transactions on Image Processing 26.4 (2017), pp. 1579–1594.

6. Huo, G., Wu, Z., Li, J., Underwater image restoration based on color correction and red channel prior, 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Miyazaki, Japan, 2018, pp. 3975–3980, https://doi.org/10.1109/SMC.2018.00674.

7. Hitam, M. S., Yussof, W. N. J. H. W., Awalludin, E. A., Bachok, Z.: Mixture contrast limited adaptive histogram equalization for underwater image enhancement. In: IEEE International Conference on Computer Applications Technology (ICCAT). 2013, pp. 1–5.

8. Zheng, H., Sun, X., Zheng, B., Nian, R., Wang, Y: Underwater image segmentation via dark channel prior and multiscale hierarchical decomposition, OCEANS 2015—Genova, Genova, Italy, 2015, pp. 1–4, https://doi.org/10.1109/OCEANS-Genova.2015.7271450

9. F. L. Pereira, T. Grilo and S. Gama (2018) Optimal Control Framework for AUV's Motion Planning in Planar Vortices Vector Field, 2018 IEEE/OES Autonomous Underwater Vehicle Workshop (AUV), Porto, Portugal, pp. 1-6, doi: https://doi.org/10.1109/AUV.2018.8729782

10. Kaeli, J. W.: Real-time anomaly detection in side-scan sonar imagery for adaptive AUV missions, 2016 IEEE/OES Autonomous Underwater Vehicles (AUV), Tokyo, Japan, 2016, pp. 85-89, doi: https://doi.org/10.1109/AUV.2016.7778653

11. Stephan, Thomas & Beyerer, Jürgen. Computergraphical Model for Underwater Image Simulation and Restoration. Proceedings - 2014 ICPR Workshop on Computer Vision for Analysis of Underwater Imagery, CVAUI 2014, 73–78. https://doi.org/10.1109/CVAUI.2014.11.

12. King, P., Vardy, A., Vandrish, P., Anstey, B.: Real-time side scan image generation and registration framework for AUV route following, 2012 IEEE/OES Autonomous Underwater Vehicles (AUV), Southampton, UK, 2012, pp. 1–6, doi: https://doi.org/10.1109/AUV.2012.6380758

13. Jiji, A. C., Nagaraj, R.: Enhancement of Underwater Deblurred Images using Gradient Guided Filter, 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 2018, pp. 1136–1140, https://doi.org/10.1109/RTEICT42901.2018.9012305.

14. Saini and M. Biswas, Object Detection in Underwater Image by Detecting Edges using Adaptive a, 2. 628–632, doi: https://doi.org/10.1109/ICOEI.2019.8862794.

15. Wang, Z., Liu, C., Wang, S., Tang, T., Tao, Y., Yang, C., Li, H., Liu, X., & Fan, X. (2020). UDD: An Underwater Open-sea Farm Object Detection Dataset for Underwater Robot Picking. ArXiv, abs/2003.01446.

16. Zhang, D., Li, L., Zhu, Z., Jin, S., Gao, W., Li, C.: Object Detection Algorithm Based on Deformable Convolutional Networks for

Underwater Images, 2019 2nd China Symposium on Cognitive Computing and Hybrid Intelligence (CCHI), Xi'an, China, 2019, pp. 274–279, doi: https://doi.org/10.1109/CCHI.2019.8901912

17. Vijayvergia, A., Kumar, K.: Selective shallow models strength integration for emotion detection using GloVe and LSTM. Multimed Tools Appl **80**, 28349–28363 (2021). https://doi.org/10.1007/s11042-021-10997-8

18. Kumar, K.: Text query based summarized event searching interface system using deep learning over cloud. Multimed Tools Appl **80**, 11079–11094 (2021). https://doi.org/10.1007/s11042-020-10157-4

19. Kumar, K. (2018). EVS-DK: Event video skimming using deep keyframe. J Vis Commun Image Represent 58. https://doi.org/10.1016/j.jvcir.2018.12.009.

20. Kumar, K., Shrimankar, D.D.: ESUMM: event SUMMarization on scale-free networks. IETE Tech. Rev. **36**(3), 265–274 (2019). https://doi.org/10.1080/02564602.2018.1454347

21. Kumar, K., Shrimankar, D.D.: F-DES: fast and deep event summarization. IEEE Trans. Multimedia **20**(2), 323–334 (2018). https://doi.org/10.1109/TMM.2017.2741423

22. Kumar, K., Shrimankar, D.D. & Singh, N. Eratosthenes sieve based key-frame extraction technique for event summarization in videos. Multimedia Tools Appl 77, 7383–7404 (2018). https://doi.org/10.1007/s11042-017-4642-9

23. J. Zhang, L. Zhu, L. Xu and Q. Xie, Research on the Correlation between Image Enhancement and Underwater Object Detection, 2020 Chinese Automation Congress (CAC), Shanghai, China, 2020, pp. 5928-5933, doi: https://doi.org/10.1109/CAC51589.2020.9326936

24. Mahmood, A., Bennamoun, M., An, S., Sohel, F., Boussaid, F., ResFeats: Residual network based features for underwater image classification, Image and Vision Computing, Volume 93, 2020, 103811, ISSN 0262-8856, https://doi.org/10.1016/j.imavis.2019.09.002

25. Cai, B., et al.: DehazeNet: an end-to-end system for single image haze removal. IEEE Trans. Image Process. **25**, 5187–5198 (2016)

26. Finlayson, G., Trezzi, E. (2004). Shades of gray and colour constancy. Proceedings of the 12th Color Imaging Conference. 37–41.

27. Huo, J. et al. Robust automatic white balance algorithm using gray color points in images. IEEE Transactions on Consumer Electronics 52 (2006): 541–546.

28. Eslami, R., Radha, H. (2005). New image transforms using hybrid wavelets and directional filter banks: analysis and design. 1. I–733. https://doi.org/10.1109/ICIP.2005.1529855.

29. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 779–788, https://doi.org/10.1109/CVPR.2016.91.

30. Szegedy, C, Ioffe, S., Vanhoucke, V., Alemi, A. A.: 2017. Inception-v4, inception-ResNet and the impact of residual connections on learning. In <i>Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence</i> (<i>AAAI'17</i>). AAAI Press, 4278–4284.

31. Han, F., Yao, J., Zhu, H., Wang, C. (2020). Underwater image processing and object detection based on deep CNN method. J. Sens. 2020.

32. Li, X., Shang, M., Hao, J., Yang, Z. Accelerating fish detection and recognition by sharing CNNs with objectness learning, OCEANS 2016—Shanghai, 2016, pp. 1–5, https://doi.org/10.1109/OCEANSAP.2016.7485476.

33. Li, X., Shang, M., Qin, H., Chen, L.: Fast accurate fish detection and recognition of underwater images with Fast R-CNN, OCEANS 2015—MTS/IEEE Washington, 2015, pp. 1–5, https://doi.org/10.23919/OCEANS.2015.7404464.