## *Underwater Change Detection Using Multiple Sampling-Based Probabilistic Learner and Feature Preservance Discriminator*

Nissar, Mehvish; Subudhi, Badri Narayan; Jakhetiya, Vinit; Mishra, Amit Kumar

*Published in:*
2024 IEEE International Conference on Image Processing

*DOI:*
10.1109/icip51287.2024.10647921

*Publication date:*
2024

*Citation for published version (APA):*
Nissar, M., Subudhi, B. N., Jakhetiya, V., & Mishra, A. K. (2024). Underwater Change Detection Using Multiple Sampling-Based Probabilistic Learner and Feature Preservance Discriminator. In *2024 IEEE International Conference on Image Processing* (pp. 3924-3930). (IEEE International Conference on Image Processing). IEEE Press. https://doi.org/10.1109/icip51287.2024.10647921

2024 IEEE International Conference on Image Processing (ICIP) | 979-8-3503-4939-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICIP51287.2024.10647921

# UNDERWATER CHANGE DETECTION USING MULTIPLE SAMPLING-BASED PROBABILISTIC LEARNER AND FEATURE PRESERVANCE DISCRIMINATOR

*Mehvish Nissar*, Badri Narayan Subudhi*, Vinit Jakhetiya*, Amit Kumar Mishra†*

*Indian Institute of Technology Jammu, NH44, Jagti, Jammu, Jammu and Kashmir, India
†Aberystwyth University, Penglais, Aberystwyth SY23 3FL, United Kingdom

## ABSTRACT

Surveillance can be defined as the process of monitoring the behavior and activities of different objects to generate meaningful insights into a video scene. In the context of underwater, surveillance can be elucidated as one of the processes of detecting and tracking the moving objects present in underwater videos. Many methods have been put forth to separate moving objects from underwater environments. Nevertheless, such methods cannot maintain the minute details that are crucial for determining an object's boundary. This is mainly due to the intricate natural properties of water and some of its characteristics, such as excessive turbidity, scattering, low visibility, etc. In this regard, we put forth an adversarial learning-based end-to-end deep learning architecture to detect underwater moving objects. The proposed architecture uses two modules for underwater object detection. The initial module is a generator comprised of a probabilistic learner which is based on multiple down-sampling and up-sampling modules. Further, the discriminator network is composed of a multi-level feature concatenation module which can perpetuate specifics at distinct levels. The effectiveness of the proposed method is confirmed using two underwater benchmark datasets by contrasting its outcomes with those of eight state-of-the-art methods.

***Index Terms—*** Underwater object detection, deep learning, adversarial learning, multi-level feature-preserving.

## 1. INTRODUCTION

Underwater surveillance focuses on the detection and tracking of objects of interest for a range of purposes, including ocean exploration, ocean life research, aquatic creature monitoring, submarine detection, etc. Since, water covers a major portion of the earth's surface, studying water ecology is one of the most important tasks, which makes underwater surveillance of great interest. The underwater images are greatly affected by many factors including the salinity of water, the number of pollutants, the size of intrinsic particles, and the scattering phenomenon. The source of illumination in underwater scenarios is either natural or artificial. The light impacting the water's surface bends in the direction of normal, and the deflected rays encounter the phenomena of reflection when they strike an object. Such reflected photons travel through water to reach the image apparatus. The particles suspended in the water cause the rays to scatter and absorb, creating a haze inside the scene. In addition to this, the reflected rays while traversing through a water medium to reach the camera device create scattering effects. The primary reason for this is the non-uniformity of the refractive index, which is brought on by variations in the water's salinity-like characteristics and temperature. Additionally, since water has distinct viscosity attenuation of optical radiation causes an image's color to degrade [1]. Conditions like poor visibility and decolorization inside aquatic environments prompt the degradation of image quality. Moreover, poor contrast issues in an image arise from the presence of all the aforementioned conditions within the water, which causes a reduction in the quantity of optical energy perceived by the camera and an increase in background intensity. Therefore, in scenarios where the color of an object is not good, and the scene view is not perspicuous, it becomes difficult to identify objects that are lying deep inside the water [2]. Furthermore, to all the above-mentioned challenges that arise in underwater scenarios, specific other ambiances exist inside water, which include occlusion, deformation, and camouflage of objects [3].

Identifying the regions in a video that create a dynamic variation in a video scene is called moving object detection. The said problem can also be quoted as the task of labeling the pixels into foreground and background. In this process, moving and non-moving objects are considered foreground and background pixels, respectively. Detecting an object of interest in underwater is a very critical and challenging task due to complex scene dynamics. Hence, it is considered a very crucial problem in the field of computer vision [4]. Therefore, there is a dire need to develop a robust object detection technique capable of handling scene dynamism [5]. Deep learning is an increasingly important field that has seen tremendous success lately. As a result, it has gained a lot of popularity in a variety of applications, including object detection [6]. To extract the essential and valuable feature representation from data [7] convolutional neural networks (CNNs) have played a remarkable role. Such neural networks are influential and dominant in extracting the features from images at various levels. In other words, CNNs are very powerful in identifying coarse, moderate, and extravagant features from images. So,

this aspect of CNN is very useful in solving various problems of computer vision domain like foreground-background segmentation where each pixel in an image needs to be assigned a class label, and such kind of prediction requires the understanding of contextual information in the scene both at higher as well as at lower level. Further, many CNN-based architectures used for underwater object detection are affected by poor distinction on accurate object boundaries.

In this article, we propose a simple and proficient end-to-end adversarial learning-based architecture to detect moving objects in underwater scenarios. Two sub-modules, named probabilistic learner and multi-level feature preserving discriminator, are included in these types of networks and are trained against one another. Imperatively, the probabilistic learner cannot directly access actual real images; instead, it can only learn by interacting with a discriminator that has access to both real and learner-generated images. The probabilistic learner is a generator module comprised of multiple down-sampling and up-sampling modules. While the multi-level feature-preserving discriminator is a discriminator network composed of component which concatenates features at multiple distinct scales. The performance of the same is evaluated on two bench-mark underwater databases. It is further verified by comparing its results with those of the eight state-of-the-art object detection techniques. The advantages of the proposed architecture are listed below:

- The use of a multi-level feature-preserving discriminator helps in defining the accurate structure of an object present in underwater videos.
- The proposed network is trained on just 5% of images yielding satisfactory results. Thus, alleviating the need to use huge amounts of data for training the model.

The rest of this article is enumerated as follows. Section 2 outlines various state-of-the-art moving object detection approaches. The strategy of the proposed methodology is briefly discussed in Section 3. The qualitative and quantitative analysis of the proposed work is depicted in Section 4. Lastly, Section 5 yields the notion of the conclusion of this article.

## 2. STATE-OF-THE-ART TECHNIQUES

The study of marine ecology and exploration has made use of underwater moving object detection techniques one of the important problems in this decade. Numerous methods, based on both non-deep learning and deep learning, have been popularly explored in the underwater domain to identify moving objects.

### 2.1. Non-deep learning based Techniques

The traditional underwater object detection techniques follow either parametric/non-parametric background modeling or a saliency-based object segmentation process.

#### *2.1.1. Parametric/Non-parametric Background Modeling Techniques*

In fixed camera-captured sequences, any change in intensity is assumed to occur because of moving objects. Most of the traditional background subtraction scheme uses multiple frames to model background scene where parameters of the model are estimated using either parametric or non-parametric estimation techniques. Further modeled background is compared against the target frames to detect the moving objects. One such landmark work is based on the Gaussian Mixture model (GMM), Stauffer *et al.* [8], where the said authors have suggested an adaptive GMM to model the dynamic background from a video. Radolko *et al.* [9] proposed a method to model spatial relationships among pixels, using a combination of background subtraction algorithm and Markov random field (MRF). The enhanced segmentation results are computed by using the belief propagation algorithm. Zivkovic *et al.* [10] introduced an improved version of adaptive GMM by using the Gaussian mixture probability density function to perform background subtraction tasks.

An updated version of the adaptive background mixture model is proposed in [11] to model background, which can distinguish between moving shadows and moving objects. In another work, Zivkovic *et al.* [12] updated the parameters of the Gaussian Mixture model by using some equations. Also, the authors have tried to improve the non-parametric methods to select the number of components per pixel and perform background subtraction. To carry out the process of change detection, Radolko *et al.* [13] proposed a method by combining flux tensor pre-segmentation with an extended Gaussian switch model.

#### *2.1.2. Saliency based Techniques*

The visually distinguishable and meaningful regions corresponding to salient regions of a video frame are also explored for moving object detection. Such algorithms aim to segment the moving object regions present in a video by considering the saliency regions in a frame. Lee *et al.* [14] used the well-known technique called Scale-invariant feature transform to capture prominent features of the object regions in a frame and further localized a correspondence between the feature points to detect moving objects in the subsequent frames. Ghosh *et al.* [15] have introduced a fuzzy edge scheme encompassing an MRF model to assign labels to images. They have explored the histogram matching placed on the chi-square test to track the deformable shaped objects of interest. Walther *et al.* [16] provided a method for underwater moving object detection based on the unique saliency of distinct color planes, which employs a bottom-up attention approach. Kumar *et al.* [17] proposed a technique to identify the in and out of objects present in the scene. The said method is motivated by the fact that the saliency of an object present in the scene does not change with the surrounding environment. Zhu *et al.* [18] introduced a saliency detection-based diver identification technique for sonar pictures.

### 2.2. Deep learning based Techniques

Recently, deep learning has continued to gain popularity owing to its ability to provide good accuracies on unstructured datasets. Li *et al.* [19] explored a fish identification system based on Faster R-CNN to detect different fish types from un-

derwater images. This method paved the way for marine biologists to understand the geography and biology of the oceanic environment. Semantic object segmentation uses well-known deep learning-based architectures for underwater surveillance one of which is U-Net. Bajpai *et al.* [6] suggested, an architecture for underwater object detection using a modified version of the U-Net in which ResNet-based encoders were used. Wang *et al.* [20] introduced YOLO nano-architecture to detect different marine species, e.g.; holothurian, echinus, starfish, etc., from underwater images.

In order to accomplish foreground-background segmentation, Lim *et.al* [21] introduced an encoder-decoder architecture based on VGG-16 to carry out the change detection procedure from complex above-water scenes. Panda *et al.* [22] have proposed an end-to-end encoder-decoder network based on VGG-19 transfer learning. In another work done in [23] the authors introduced an encoder-decoder architecture based on ResNet-50 to carry out the change detection procedure from complex above-water scenes, by reducing the amount of trainable parameters. Later, in [24] the authors proposed a modified ResNet-152-based encoder-decoder background subtraction model to segment moving objects from complicated outdoor situations.

## 3. PROPOSED METHOD

This endeavor aims to construct a network that can identify moving objects from underwater scenarios. Because dynamics in an aquatic environment are different from those in terrestrial domains, finding an object in an aquatic context is more difficult than recognizing it in a terrestrial scenario. Also, underwater scenes provide a variety of difficulties, such as haze, decolorization, low contrast, poor lighting, object deformation, etc. Thus, our work suggests an architecture that helps to detect objects from underwater situations by preserving high-level and low-level contextual information. The proposed network is built on the modified Generative Adversarial Networks (GAN) architecture, which produces the required images after learning the data distribution. We have tried to embed the module inside the discriminator, which is capable of extracting low, mid, and high-level features from an image. Accordingly, the generator is trained to learn better data distribution to segment the interested object from underwater images. Our proposed architecture has two modules, namely- Probabilistic Learner and Multi-level feature preserving discriminator, respectively, as depicted in Fig. 1.

### 3.1. Probabilistic Learner

As depicted in Fig. 1, the probabilistic learner module possesses several up-sampling and down-sampling blocks to extract hierarchical characteristics from the given underwater video frame. A series of down-sampling operations are initially performed on the input image to extricate relevant characteristics or information while lowering the spatial dimensions. A sequence of up-sampling procedures is applied to restore the spatial dimensions. These up-sampling techniques integrate features from analogous down-sampling operations through skip connections. There are two blocks in this module termed as Down-Sampling Block and Up-Sampling Block respectively. The Down-Sampling Block comprises eight smaller sub-blocks (Sb), each carrying out a particular set of operations: convolution process, batch normalization, and leaky rectified linear unit in that order. While the number of filters in each sub-block varies, the $4 \times 4$ filter size is equal for all sub-blocks. There are 64, 128, and 256 filters in the first, second, and third sub-blocks, respectively, and 512 filters in the fourth through eighth sub-blocks. The output of the first sub-block is passed as an input to the one after it, and the last sub-block's output is eventually provided as an input to the Up-Sampling Block. This block also consists of eight sub-blocks (Sb), the first seven carry out a series of transposed convolution operations before applying batch normalization, dropout, and rectified linear units in that order. Aside from these functions, every sub-block passes its output as an input to the succeeding sub-block after concatenating it with the matching sub-block in the first block. The number of filters in each sub-block varies, but the filter's $4 \times 4$ size is equal for all sub-blocks. There are 512 filters in the first through fourth sub-blocks and 256, 128, and 64 filters in the fifth, sixth, and seventh sub-blocks, respectively. The last sub-block in block two receives the output of the seventh sub-block as an input and applies a transposed convolution operation to produce the output that is the generator's output.

The discriminator's task is to measure the similarity between the actual image and the image generated by the generator. To ascertain the degree of actuality between them, the discriminator is fed with a target image and an input image, an output image created by the generator, and an input image in that order. Both pairs are considered when the discriminator computes loss and weights of the same component are updated accordingly. Subsequently, the output of the discriminator that is acquired from the generator's output, and the input image is taken into account to adjust the generator's weights. Furthermore, while estimating generator loss, the difference between the target and generator-produced images is considered.

### 3.2. Multi-level feature preserving discriminator (MLFD)

The novelty of this work is in the designing of the discriminator part based on a multi-level feature concatenation component (MFC) which employs several layers of max pooling and dilation convolution blocks. As shown in Fig. 1, there are two blocks in this module as well. It concatenates the input image, starting with the target image and moving upon the generated image in that order. Every pair goes through blocks one and two separately. The first block termed MFC, is responsible for combining features at various scale, preserving minute and significant object details. At high-level and low-level features it is very crucial to lose contextual information while we are doing pixel labeling tasks. In addition to it, for semantic segmentation issues, high-resolution feature maps must be captured. To retain such information MFC

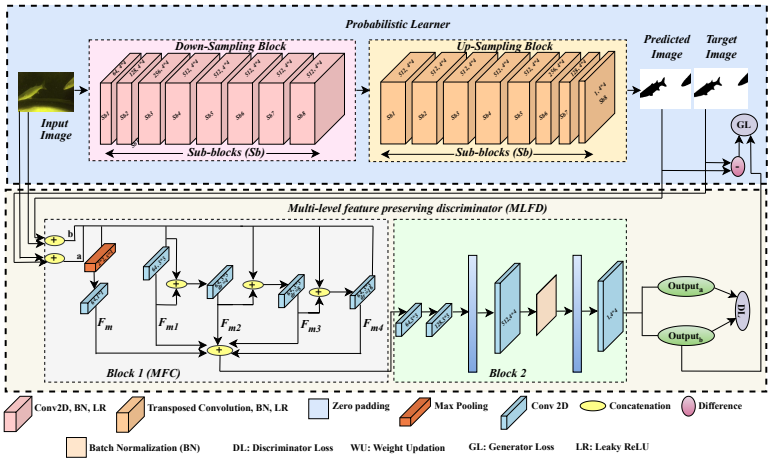


**Fig. 1**: Proposed architecture for underwater object detection.

block plays a vital role. Each concatenated input is first run through $2 \times 2$ max pooling with a stride of 1 and then a convolution layer with 64 filters each of size $1 \times 1$. This function yields a feature map, denoted as $Fm$. Once again, the input is sent through a convolution layer of 64 filters, each measuring $3 \times 3$, yielding $Fm_1$. The same outcome is combined with input, and run again through a convolution layer comprising 64 filters each $3 \times 3$ in size with a dilation rate of 4, and then an activation function known as a rectified linear unit (ReLU) is used, which results in $Fm_2$. Following another concatenation of input with $Fm_2$, a convolution layer possessing 64 filters each of size $3 \times 3$, with an 8-dilation rate is applied. Then, the rectified linear unit (ReLU) activation function is used, endowing $Fm_3$ as a result. Afterward, the input is once more combined with $Fm_3$, and a convolution layer having 64 filters measuring $3 \times 3$ with a dilation rate of 16 is practiced, and the ReLU activation function is then applied, providing $Fm_4$ as output. Following all of these procedures, Fm, $Fm_1$, $Fm_2$, $Fm_3$, $Fm_4$ are concatenated along the depth dimension, which comprises characteristics along multiple levels. The final concatenated feature map is subsequently put through the ReLU activation function and spatial dropout. The first block's output feeds into the second block, which is run through a convolution layer with 64 filters, each measuring $3 \times 3$, followed by using a rectified linear unit (ReLU) activation function. This layer's output is sent into a second convolution layer with 128 filters, each of size $3 \times 3$, and then the ReLU activation function is applied. Following these operations, we applied zero padding, a convolution layer with 512 filters, each of size $4 \times 4$, and batch normalization in that sequence. Again, zero padding and a convolution layer but with a single, $4 \times 4$ filter is applied. The discriminator's action is performed by this second module, which updates its weights and computes the loss by accounting for the previously stated inputs. For calculating the loss, the generator considers the discriminator's output derived from the concatenation of the input image with its output. Also, the estimation of the same loss considers the difference between the output produced by the generator and the actual output. In this manner, the generator adjusts its weights and progressively gains a better distribution of data, resulting in the production of an output that contains segmented objects of interest.

The proposed method is executed using Python programming and NVIDIA A100 80GB GPU with 256GB RAM. We have assessed the performance of the proposed scheme on a standard Underwater change detection [25] and Fish4Knowledge dataset [26] respectively. The first one comprises of five challenges: caustics, fish swarm, small aquaculture, marine snow, and two fishes. In the second one, we have considered six challenges: complex background 1, complex background 2, crowded, hybrid, standard, and lumi-change.

### 3.3. Model Parameters

We have used Adam, as our optimizer because we wish to avoid advancing more quickly or leaping over minima. Instead, we would like to slow down the pace slightly to get better search and practical model training. Epsilon and beta are the hyper-parameters that we have applied to Adam. To determine how many previous gradients should be taken into account while updating the parameters, beta is used. For training the model, we have set the beta value equal to 0.5. To prevent division by zero circumstance, another hyper-parameter known as epsilon is used which is a constant, added to the denominator to avoid the same issue from arising for the optimizer. $2e^{-4}$ is the epsilon value we have chosen for our model. The model is trained over 70 epochs for every category by giving details about the input image and their corresponding output. Only 5% of the training images from each category in the Underwater change detection and Fish4Knowledge dataset are used to train our proposed net-

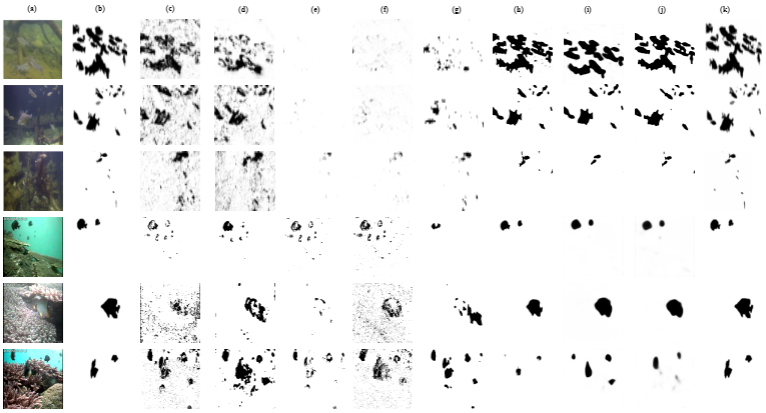

**Fig. 2**: Visual Results: (a) original image (b) ground-truth, results obtained by (c) GSMM [9], (d) AGMM [10], (e) ABMM [11], (f) ADE [12], (g) GWFT [13], (h) GAFP [21], (i) MFPD [22], (j) HPPM [24] and (k) PM.

**Table 1**: Quantitative analysis of Underwater change detection dataset in terms of F-measure with eight SOTA architectures.

| Dataset Categories | GSMM[9] | AGMM[10] | ABMM[11] | ADE[12] | GWFT[13] | GAFP [21] | MFPD [22] | HPPM [24] | PM |
|---|---|---|---|---|---|---|---|---|---|
| Marine Snow | 0.84 | 0.82 | 0.65 | 0.82 | 0.91 | 0.96 | 0.95 | 0.92 | 0.98 |
| Small Aquaculture | 0.77 | 0.74 | 0.43 | 0.88 | 0.93 | 0.88 | 0.86 | 0.90 | 0.98 |
| Two Fishes | 0.79 | 0.79 | 0.76 | 0.71 | 0.82 | 0.96 | 0.94 | 0.91 | 0.99 |
| Caustics | 0.55 | 0.74 | 0.67 | 0.75 | 0.67 | 0.97 | 0.91 | 0.96 | 0.99 |
| Fish Swarm | 0.57 | 0.30 | 0.06 | 0.59 | 0.85 | 0.92 | 0.75 | 0.91 | 0.99 |

**Table 2**: Quantitative analysis of Fish4Knowledge dataset in terms of F-measure with eight SOTA architectures.

| Dataset Categories | GSMM[9] | AGMM[10] | ABMM[11] | ADE[12] | GWFT[13] | GAFP [21] | MFPD [22] | HPPM [24] | PM |
|---|---|---|---|---|---|---|---|---|---|
| Complex Background 1 | 0.30 | 0.25 | 0.26 | 0.13 | 0.42 | 0.51 | 0.20 | 0.16 | 0.99 |
| Complex Bcakground 2 | 0.19 | 0.50 | 0.21 | 0.16 | 0.32 | 0.77 | 0.17 | 0.52 | 0.99 |
| Crowded | 0.42 | 0.57 | 0.40 | 0.30 | 0.51 | 0.65 | 0.13 | 0.54 | 0.98 |
| Hybrid | 0.30 | 0.26 | 0.35 | 0.48 | 0.24 | 0.84 | 0.15 | 0.22 | 0.98 |
| Standard | 0.76 | 0.64 | 0.75 | 0.48 | 0.64 | 0.81 | 0.55 | 0.39 | 0.99 |
| LumiChange | 0.72 | 0.73 | 0.54 | 0.76 | 0.16 | 0.85 | 0.14 | 0.58 | 0.99 |

**Table 3**: Quantitative analysis of proposed method in terms of Precision, Recall, and F-measure on different challenges of Underwater change detection dataset.

| Dataset Categories | Precision | Recall | F-measure |
|---|---|---|---|
| Marine Snow | 0.99 | 0.98 | 0.98 |
| Small Aquaculture | 0.99 | 0.97 | 0.98 |
| Two Fishes | 0.99 | 0.99 | 0.99 |
| Caustics | 0.99 | 0.99 | 0.99 |
| Fish Swarm | 0.98 | 0.97 | 0.99 |

**Table 4**: Quantitative analysis of proposed method in terms of precision, recall, and F-measure on different challenges of Fish4Knowledge dataset.

| Dataset Categories | Precision | Recall | F-measure |
|---|---|---|---|
| Complex Background | 0.99 | 0.99 | 0.99 |
| Complex Background 2 | 0.98 | 0.99 | 0.99 |
| Crowded | 0.99 | 0.98 | 0.98 |
| Hybrid | 0.97 | 0.99 | 0.98 |
| Standard | 0.97 | 0.98 | 0.99 |
| LumiChange | 0.99 | 0.99 | 0.99 |

**Table 5**: Ablation study of different discriminator architectures on the Underwater change detection and Fish4Knowledge dataset.

| Methods | Average Precision | Average Recall | Average F-measure |
|---|---|---|---|
| Underwater change detection dataset | | | |
| PM | 0.99 | 0.99 | 0.99 |
| W-MLFC | 0.70 | 0.85 | 0.76 |
| C-MLFC | 0.92 | 0.86 | 0.88 |
| Fish4Knowledge dataset | | | |
| PM | 0.99 | 0.99 | 0.99 |
| W-MLFC | 0.60 | 0.71 | 0.65 |
| C-MLFC | 0.67 | 0.78 | 0.72 |

**Table 6**: Quantitative study on parameter comparison.

| Methods | Total Parameters | Trainable Parameters | Non-Trainable Parameters |
|---|---|---|---|
| GAFP [21] | 7,635,264 | 5,899,776 | 1,735,488 |
| MFPD [22] | 10,585,152 | 8,259,584 | 2,325,568 |
| HPPM [24] | 2,673,344 | 2,339,840 | 333,504 |
| PM | 55,867,524 | 55,855,620 | 11,904 |

work. To avoid the case of over-fitting with a lesser amount of data, L2 regularization is used.

### 3.4. Qualitative Evaluation

The proposed architecture is qualitatively assessed using different challenges of the Underwater change detection and Fish4Knowledge dataset. Fig. 2 provides the visual details of the findings on the aforementioned datasets. The Actual input and ground truth are represented in columns (a) and (b) respectively. The qualitative evaluation of the proposed scheme is carried out using eight SOTA techniques. The results obtained by those of the considered SOTA techniques: GSMM [9], AGMM [10], ABMM [11], ADE [12], GWFT [13], GAFP [21], MFPD [22], HPPM [24] are shown in Fig. 2 (c)-(j). It may be observed from these results that all the considered SOTA techniques provide results where the actual object structure is lost and many false alarms appear on the objects. The results obtained by the proposed method (PM) are shown in Fig. 2-column (k) which indicates that the pro-

posed technique provides better results as compared to the considered SOTA techniques. Also, the proposed method's outcome indicates that it can maintain the crucial information that defines an object's structure, making it suitable for object detection.

### 3.5. Quantitative Evaluation

In the proposed scheme, three evaluation metrics are used to complete the quantitative appraisal of the proposed architecture: F-measure, precision, and recall. As indicated in Table 1 and 2, the effectiveness of the proposed method is supported by comparing the results obtained by it with those of the eight state-of-the-art approaches: GSMM [9], AGMM [10], ABMM [11], ADE [12], GWFT [13], GAFP [21], MFPD [22], and HPPM [24] in terms of F-measure. When compared to other SOTA techniques on the considered datasets, it is obvious from the table that the proposed architecture is substantially outperforming in detecting moving objects from underwater sequences. While these conventional techniques may identify objects to a certain extent, they are not capable of maintaining an object's silhouette, in comparison to the proposed method. These results exhibit the efficacy of our network in object detection. It is also evident from Table 1 and 2 that the proposed strategy has produced very satisfactory results in terms of F-measure. Even the algorithm has proven to be quite effective in identifying small objects in underwater videos. The evaluation of the proposed scheme on the considered databases is carried out using precision, recall, and F-measure and is reported in Table 3 and 4. It is evident from the values shown in Table 3 and 4 that for each category our proposed network has done a quite better job of identifying or segmenting the moving objects from the underwater videos. Table 6 shows the comparison of parameters obtained with GAFP [21], MFPD [22], HPPM [24] and PM. Although, PM yields more parameters than GAFP [21], MFPD [22], and HPPM [24], but both the quantitative and visual outcomes are significantly superior to those of GAFP [21], MFPD [22] and HPPM [24].

### 3.6. Ablation Study

The ablation analysis of the proposed method is conducted on a benchmark Underwater change detection and Fish4Knowledge dataset using multiple discriminator designs. One architecture (W-MFC) uses a stack of specific convolution layers in place of the MFC in the discriminator. In another (C-MFC), there is no use of the convolution's dilation rate and no concatenation of the convolution with input at certain levels. For the comparison, the evaluation metrics: Average Recall, Average Precision, and Average F1-score are utilized. Table 5 unequivocally shows that all of the parameter values are significantly better when using the proposed method with the multi-level feature concatenation component.

## 4. CONCLUSION

This research proposed a new architecture for underwater object detection that uses a generative adversarial network by integrating a multi-level feature concatenation component in the discriminator part of the network. The proposed architecture is composed of a probabilistic learner which has multiple down and up-sampling modules. Further, a feature extraction across different scales followed by a concatenation of the same, based discriminator module is designed to preserve the structure of objects. The proposed architecture undergoes training, validation, and testing using two standard underwater datasets, each with various challenges. Using just a few training examples, our network learns the minute details of the underwater scene to identify the moving objects. Moreover, this module mitigated the need to use inputs at multiple scales for training the network. Despite requiring a substantial amount of memory space and increasing total parameters due to the introduction of the said module within the discriminator, our solution eliminates the need for any post-processing steps to enhance the final segmentation results.

## 5. ACKNOWLEDGMENTS

Mehvish Nissar acknowledges Department of Science & Technology, Government of India for awarding DST INSPIRE Fellowship bearing registration Number IF220151.

## 6. REFERENCES

[1] Risheng Liu, Xin Fan, Ming Zhu, Minjun Hou, and Zhongxuan Luo, "Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light," *IEEE transactions on circuits and systems for video technology*, vol. 30, no. 12, pp. 4861–4875, 2020.

[2] Srikanth Vasamsetti, Supriya Setia, Neerja Mittal, Harish K Sardana, and Geetanjali Babbar, "Automatic underwater moving object detection using multi-feature integration framework in complex backgrounds," *IET Computer Vision*, vol. 12, no. 6, pp. 770–778, 2018.

[3] Deepak Kumar Rout, Meghna Kapoor, Badri Narayan Subudhi, Veerakumar Thangaraj, and Vinit Jakhetiya, "Passive visual underwater surveillance: A survey," *Research Square (Preprint)*, 2023.

[4] Deepak Kumar Rout, Badri Narayan Subudhi, Thangaraj Veerakumar, and Santanu Chaudhury, "Walsh–Hadamard-kernel-based features in particle filter framework for underwater object tracking," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 9, pp. 5712–5722, 2019.

[5] Deepak Kumar Rout, Badri Narayan Subudhi, Thangaraj Veerakumar, and Santanu Chaudhury, "Spatio-contextual Gaussian mixture model for local change detection in underwater video," *Expert Systems with Applications*, vol. 97, pp. 117–136, 2018.

[6] Vatsalya Bajpai, Akhilesh Sharma, Badri Narayan Subudhi, T Veerakumar, and Vinit Jakhetiya, "Underwater U-Net: Deep learning with u-net for visual underwater

moving object detection," in *OCEANS*. IEEE, 2021, pp. 1–4.

[7] Matthew D Zeiler and Rob Fergus, "Visualizing and understanding convolutional networks," in *Proceedings of the European Conference of Computer Vision*, 2014, pp. 818–833.

[8] Chris Stauffer and W. Eric L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 8, pp. 747–757, 2000.

[9] Martin Radolko and Enrico Gutzeit, "Video segmentation via a Gaussian switch background model and higher order Markov random fields.," in *Proceedings of the International Conference on Computer Vision Theory and Applications*. SCITEPRESS, 2015, vol. 2, pp. 537–544.

[10] Zoran Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," in *Proceedings of the IEEE International Conference on Pattern Recognition*, 2004, vol. 2, pp. 28–31.

[11] Pakorn KaewTraKulPong and Richard Bowden, "An improved adaptive background mixture model for real-time tracking with shadow detection," *Video-based surveillance systems: Computer vision and distributed processing*, pp. 135–144, 2002.

[12] Zoran Zivkovic and Ferdinand Van Der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern recognition letters*, vol. 27, no. 7, pp. 773–780, 2006.

[13] Martin Radolko, Fahimeh Farhadifard, and Uwe von Lukas, "Change detection in crowded underwater scenes-via an extended Gaussian switch model combined with a flux tensor pre-segmentation," in *Proceedings of the International Conference on Computer Vision Theory and Applications*. SCITEPRESS, 2017, vol. 5, pp. 405–415.

[14] Donghwa Lee, Gonyop Kim, Donghoon Kim, Hyun Myung, and Hyun-Taek Choi, "Vision-based object detection and tracking for autonomous navigation of underwater robots," *Ocean Engineering*, vol. 48, pp. 59–68, 2012.

[15] Ashish Ghosh, Badri Narayan Subudhi, and Susmita Ghosh, "Object detection from videos captured by moving camera by fuzzy edge incorporated Markov random field and local histogram matching," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 8, pp. 1127–1135, 2012.

[16] Dirk Walther, Duane R Edgington, and Christof Koch, "Detection and tracking of objects in underwater video," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2004, vol. 1, pp. I–I.

[17] Nitin Kumar, Harish Kumar Sardana, SN Shome, and Neerja Mittal, "Saliency subtraction inspired automated event detection in underwater environments," *Cognitive Computation*, vol. 12, pp. 115–127, 2020.

[18] Jianjiang Zhu, Siquan Yu, Lei Gao, Zhi Han, and Yandong Tang, "Saliency-based diver target detection and localization method," *Mathematical Problems in Engineering*, vol. 2020, pp. 1–14, 2020.

[19] Xiu Li, Min Shang, Hongwei Qin, and Liansheng Chen, "Fast accurate fish detection and recognition of underwater images with fast R-CNN," in *OCEANS*. IEEE, 2015, pp. 1–5.

[20] Lin Wang, Xiufen Ye, Huiming Xing, Zhengyang Wang, and Peng Li, "Yolo nano underwater: A fast and compact object detector for embedded device," in *Oceans*. IEEE, 2020, pp. 1–4.

[21] Long Ang Lim and Hacer Yalim Keles, "Learning multi-scale features for foreground segmentation," *Pattern Analysis and Applications*, vol. 23, no. 3, pp. 1369–1380, 2020.

[22] Manoj Kumar Panda, Badri Narayan Subudhi, Thierry Bouwmans, Vinit Jakheytiya, and T Veerakumar, "An end to end encoder-decoder network with multi-scale feature pulling for detecting local changes from video scene," in *Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance*, 2022, pp. 1–8.

[23] Manoj Kumar Panda, Akhilesh Sharma, Vatsalya Bajpai, Badri Narayan Subudhi, Veerakumar Thangaraj, and Vinit Jakhetiya, "Encoder and decoder network with ResNet-50 and global average feature pooling for local change detection," *Computer Vision and Image Understanding*, vol. 222, pp. 103501, 2022.

[24] Manoj Kumar Panda, Badri Narayan Subudhi, Thangaraj Veerakumar, and Vinit Jakhetiya, "Modified ResNet-152 network with hybrid pyramidal pooling for local change detection," *IEEE Transactions on Artificial Intelligence*, 2023.

[25] Martin Radolko, Fahimeh Farhadifard, and Uwe Freiherr von Lukas, "Dataset on underwater change detection," in *OCEANS*. IEEE, 2016, pp. 1–8.

[26] Robert B Fisher, Yun-Heh Chen-Burger, Daniela Giordano, Lynda Hardman, Fang-Pang Lin, et al., *Fish4Knowledge: collecting and analyzing massive coral reef fish video data*, vol. 104, Springer, 2016.