



OPEN ACCESS

EDITED BY

Hong Song,
Zhejiang University, China

REVIEWED BY

Muhammad Abdul Rehman Khan,
Air University, Pakistan
Muhammad Basim,
Sungkyunkwan University, South Korea
Danish Khan,
Zhejiang University, China
Muwei Jian,
Shandong University of Finance and
Economics, China

*CORRESPONDENCE

Taejoon Kim

✉ ktjcc@chungbuk.ac.kr

Inam Ullah

✉ inam.fragrance@gmail.com

[†]These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION

This article was submitted to Ocean Observation, a section of the journal Frontiers in Marine Science

RECEIVED 10 November 2022

ACCEPTED 19 December 2022

PUBLISHED 11 January 2023

CITATION

Khan S, Ullah I, Ali F, Shafiq M, Ghadi YY and Kim T (2023) Deep learning-based marine big data fusion for ocean environment monitoring: Towards shape optimization and salient objects detection. *Front. Mar. Sci.* 9:1094915. doi: 10.3389/fmars.2022.1094915

COPYRIGHT

© 2023 Khan, Ullah, Ali, Shafiq, Ghadi and Kim. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Deep learning-based marine big data fusion for ocean environment monitoring: Towards shape optimization and salient objects detection

Sulaiman Khan^{1†}, Inam Ullah^{2*†}, Farhad Ali³, Muhammad Shafiq^{4,5}, Yazeed Yasin Ghadi⁶ and Taejoon Kim^{7*}

¹College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar,

²BK21 Chungbuk Information Technology Education and Research Center, Chungbuk National University, Cheongju, Republic of Korea, ³Department of Accounting and Information Systems,

College of Business and Economics, Qatar University, Doha, Qatar, ⁴Cyberspace Institute of

Advanced Technology, Guangzhou University, Guangzhou China, ⁵Department of Computer

Science, Adjunct Professor at Shenyang Normal University, Shenyang, China, ⁶Department of

Computer Science, Al Ain University, Abu Dhabi, United Arab Emirates, ⁷School of Information and

Communication Engineering, Chungbuk National University, Cheongju, Republic of Korea

Objective: During the last few years, underwater object detection and marine resource utilization have gained significant attention from researchers and become active research hotspots in underwater image processing and analysis domains. This research study presents a data fusion-based method for underwater salient object detection and ocean environment monitoring by utilizing a deep model.

Methodology: A hybrid model consists of an upgraded AlexNet with Inception v-4 for salient object detection and ocean environment monitoring. For the categorization of spatial data, AlexNet is utilized, whereas Inception V-4 is employed for temporal data (environment monitoring). Moreover, we used preprocessing techniques before the classification task for underwater image enhancement, segmentation, noise and fog removal, restoration, and color constancy.

Conclusion: The Real-Time Underwater Image Enhancement (RUIE) dataset and the Marine Underwater Environment Database (MUED) dataset are used in this research project's data fusion and experimental activities, respectively. Root mean square error (RMSE), computing usage, and accuracy are used to

construct the model's simulation results. The suggested model's relevance form optimization and conspicuous item prediction issues in the seas is illustrated by the greatest accuracy of 95.7% and low RMSE value of 49 when compared to other baseline models.

KEYWORDS

data fusion, marine big data, ocean environment, underwater saliency detection, underwater image processing

1 Introduction

Oceans cover about 70 percent of the earth's surface and hold enriched natural resources that endeavor in long-term human development. With the emergence of advanced information and communication systems, the researchers paid significant attention to the exploration of mysterious areas of the ocean, marine exploitation, and observation. However, facing relatively hashed and unconstrained marine scenes, there are legion austere inimical factors such as water with high turbidity, dreary color, uneven illustration, and a perilous underwater atmosphere that gravely compromises the accuracy and accessibility of underwater photos in real-world contexts (Jian et al., 2021). To tackle these issues, numerous research reports have been published on how to effectively perform underwater image and visionary processing tasks. Qiao et al. (2022), proposed a novel generative adversarial network (GAN) architecture to address color distortion or under/overexposure problems in underwater images. Their model generates high-quality enhanced images to address the non-uniform illumination problems in images. Moghimi and Mohanna systematically analyzed the extant data and identified different real-time underwater image enhancement models proposed in the literature (Moghimi and Mohanna, 2021). Halimi et al. present an effective technique for boosting underwater image resolution by concurrently reconstructing the reflectivity and complexity of the image using the greatest marginal probability estimate (Halimi et al., 2017).

The aim of salient object recognition, which makes use of image/video segmentation, is to pinpoint the most eye-catching and aesthetically attractive objects or regions in a picture (Wang et al., 2018), image foreground annotation (Cao et al., 2016), image quality assessment (Gu et al., 2016), and video summarization (Cong et al., 2019). In-depth research has been carried out utilizing very superior artificial intelligence (AI) algorithms to identify saliency in photographs of natural settings during the past few years. It is imperative to note that salient object identification differs greatly from studies on anomaly detection and conventional object detection. First of all, unlike anomaly detection and salient object recognition,

which only commit to discovering locally notable objects, object detection is a comprehensive job that focuses on detecting everything. Salient object detection produces a saliency probability map at the pixel level as opposed to object detection and anomaly detection, which always identify items with bounding boxes. The researchers used a range of algorithms to discover significant details in images and videos, such as Lee et al. (2018), proposal of a manipulated deep learning model for saliency detection using a convolutional neural network (CNN) and GoogLeNet. This model uses both low-level and high-level features. Google is used to gather high-level data, while CNN architecture is used to extract low-level attributes. By comparing a local region's differences from other areas in an image, these attributes were utilized to assess a region's relative importance. Zhang et al. (2019), proposed a pipelined model for detecting salient objects in underwater photos, using deformable convolutional networks. Using the CNN approach, scientists first reduced noise in underwater photographs to improve contrast. Increase the feature extraction capabilities by implementing a deformable position-sensitive ROI pooling method with RPN and rebuilding the ResNet-101 feature extraction sub-network by using a deformed convolution model.

After studying the literature, it was concluded that most of the models either had a low identification rate, especially in high turbidities, or were highly time-consuming and required more data for the testing and validation process. Our research work addresses these problems with the following key objectives:

- To develop a hybrid deep learning model consists of upgraded AlexNet with Inception v-4 for salient object detection and ocean environment monitoring. AlexNet is considered for the spatial data classification while Inception v-4 is exploited for temporal data (environment monitoring).
- To perform semantic-based data fusion of two publicly available datasets (MUED and RUIE) and perform training and validation tasks of the hybrid model.
- To accurately identify saliencies and perform shape optimization using this hybrid model and fused database. An overall identification rate of 95.7% is

achieved by this model. After comparing its performance with benchmark techniques, our model outperformed even on smaller dataset.

The remainder of this research article is organized by introducing the representative models on different underwater image processing and data fusion-based models in Section 2. Section 3 introduces the experimental setup exploited to accomplish this research work. Results and discussion are outlined in Section 4 of the paper followed by the conclusion in Section 5.

2 Literature review

This portion of the study summarizes recent research reports in the subject area under consideration and several data fusion-based models that have been proposed by academics for deployment in various deep learning application domains.

2.1 Underwater salient objects detection

The congenital capabilities of humans to percept, efficiently perceive, and distinguish salient objects in videos or images keep them ahead of machines (Li et al., 2019). Saliency detection is used in the machine vision and underwater image processing sectors to provide computers the capacity of people to analyze underwater images, which also plays a key role in discovering marine resources. Consequently, many underwater saliency detection models have been developed freshly.

The problem of identifying the prominent components from underwater photos using conventional saliency detection techniques is made more difficult by the uncertainty of underwater settings. The intricacy and diversity of undersea ecosystems have recently sparked a lot of study attention. Xu et al. (2019), proposed Generalized robust principal component analysis (GRPCA) as a unique method for underwater target detection. This model functions by extracting visual characteristics from underwater pictures that have specific goal in recognition and images representation. –, developed an effective underwater target layered background framework based on the visual information perception and processing of a frog's eye, which can distinguish prominent objects from the background of a picture with object contour. Chen et al. (2017), suggested employing a monocular vision sensor to recognize underwater objects using a prominent object identification approach. This technique reduced background noise in order to increase underwater detection accuracy (Ullah et al., 2019; Khan et al., 2019; Su et al., 2020). Jian et al. (Jian et al., 2021; Jian et al., 2021), suggested forward and backward cues for saliency objects detection using visual spatial temporal features. For object localization and tracking they defined a weighted centroid calculation algorithm for center prior generation and

tracked it in scene images. A real-time neural network architecture is used for saliencies detection in frames.

2.2 Data fusion background

Data fusion has drawn a lot of interest in data mining and is being used in a variety of research fields, including both civilian and military applications including monitoring animal habitats, risks identification, surveillance, and espionage operations. Data fusion techniques were initially employed to combine data into a single, featured dataset that corresponded to the converted matrix (Maragos et al., 2008). The fusion process is now carried out through data fusion techniques, which collect data characteristics from several sources. For instance, the most basic type of data fusion is merging two one-dimensional datasets. Additionally, it may be done by combining the semantics of the data. In accordance with the characteristics of the data and the machine learning approach employed for the fusion process, many data fusion algorithms produce different optimization solutions. Zheng et al. (Zheng, 2015), categorized data fusion methods into three different classes (feature-based fusion, semantic-based fusion, and stage-based fusion), as shown in Figure 1.

When adopting feature-based data fusion, the same dimensional data characteristics are directly combined after being collected from various sources, and then they are evaluated using machine learning models. Before the direct catenation of data features, careful consideration is given to the elimination of duplicate records before the merging process; second, keep the same dimensions for all the records in the fused database because model performance can be adversely affected during the merging process if you do not; and third, over-fitting is another significant challenge to face during the training process. In the stage-based data fusion models, the characteristics of the data are classified into several classes, and the data from each class is subsequently examined and combined appropriately (Zhu et al., 2018). The foundation of the semantic meaning-based data fusion method is the semantics of the data. The technique of semantic-based data fusion divides data semantics into four categories: transfer learning data fusion, multi-view data fusion, similarity-based data fusion, and probabilistic dependency data fusion. Using data derived from several sources, data values are obtained and assessed in a multi-view-based data fusion system. Co-training, multi-kernel learning, and subspace learning procedures are subcategories of this fusion process. Khan et al. (2021), performed semantic-based data fusion for traffic monitoring and flow predictions in smart cities.

3 Experimental setup

As shown in Figure 2, this study used a hybrid deep learning model made up of Inception v-4 and AlexNet. The AlexNet and

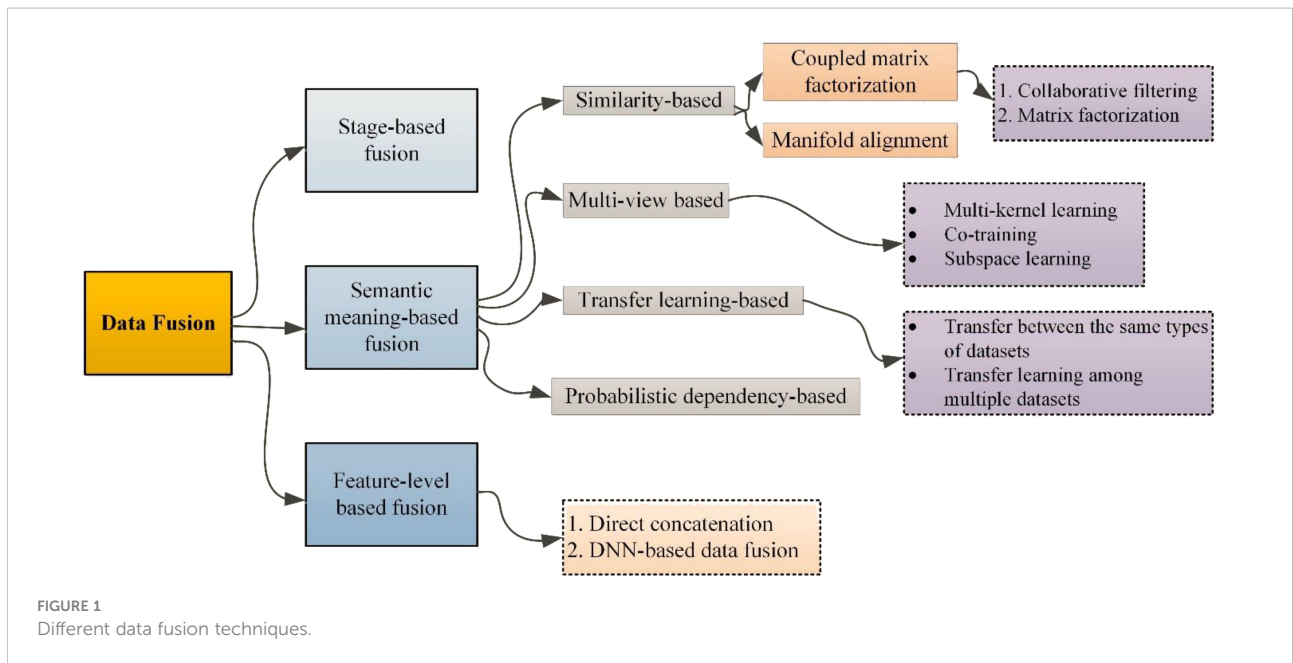


FIGURE 1 Different data fusion techniques.

inception-v4 architecture are merged in this research project. The revised AlexNet structure and the addition of the Inception-v4 module increase the network’s reprocessing capacity. The batch normalization layer (BN) is also used to increase generalization abilities, hasten convergence, and stop the gradient from fading. During the training process, mean μ and variance σ is calculated using (Eq.1) and (Eq.2) to normalize each sample in the batch. The normalized bath values are calculated using (Eq.3).

$$\mu = \frac{1}{N} \sum_{i=1}^k x_i \tag{1}$$

$$\mu = \frac{1}{N} \sum_{i=1}^k (x_i - \mu)^2 \tag{2}$$

$$\hat{x}_i = \frac{(x_i - \mu)}{\sqrt{(\sigma^2 - \epsilon)}} \tag{3}$$

where N represents total number of samples in a batch and x_i represents the input samples. The BN layer’s convolution output for a two-dimensional image input is (N, C, W, H), where W and H stand for the dimension of the feature map and C stands for the number of output channels. Afterward, each batch sample may be expressed separately as $x_{c, w, h}$. Each sample is individually normalized by the BN, thus the resultant number of μ is also $C \times W \times H$.

The BN layer has the ability to manage gradient explosion, restrict the gradient from vanishing, inhibit overfitting, and accelerate network training and convergence. Figure 2 shows how the inception-v4 model’s structure was employed in the

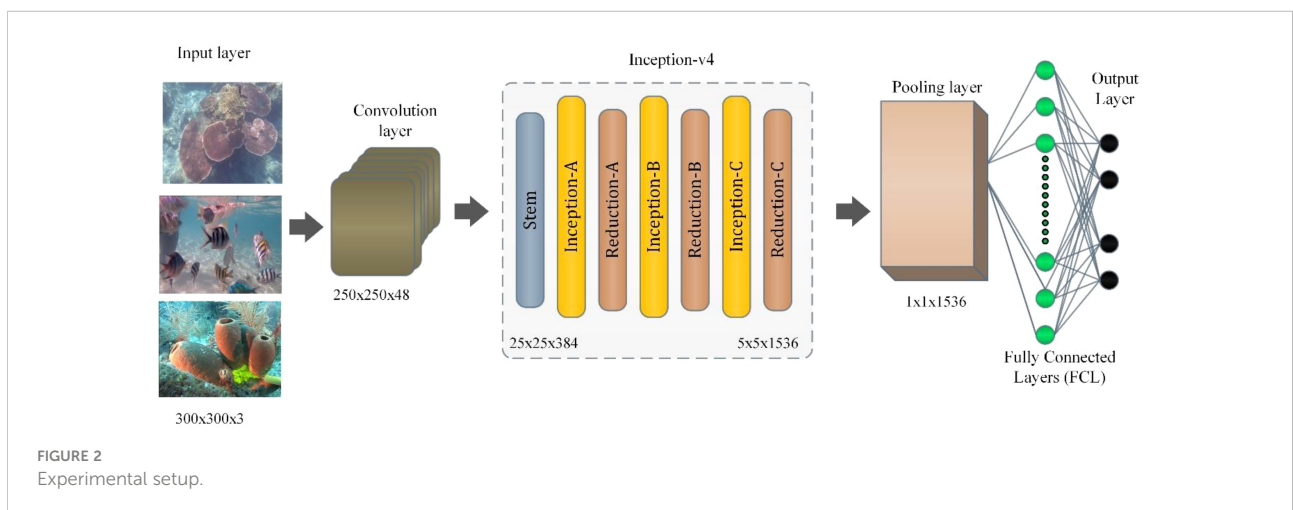


FIGURE 2 Experimental setup.

proposed research project. Input data (images) are provided at the input layer. The number of hidden layers can be defined based on the nature of the research problem and hardware specifications. A huge amount of data and many hidden layers require more simulation time for a simple hardware combination (normal computers with no GPU installed). The output layers, or fully connected layers (FCL), generate the output based on the information generated from hidden layers. It can also be termed the “SoftMax layer.” Different activation functions like rectified linear unit (ReLU) and tanh are used for validation purposes. This research work uses transfer learning with five hidden layers. The convolution layer calculates texture information from the input image. Following then, the Inception-V4 continues to feature extraction as a communications system, which is made up of several convolutional and pooling procedures.

The Inception-X module, where X stands for A, B, and C, improves feature utilization by teaching image features through multiple concurrent feature transferring mechanisms. Reduction-X shrinks large feature maps into small feature maps while increasing the number of channels, with X denoting the A and B modules. This research work proposes monocular vision sensor to recognize underwater objects using a prominent object identification approach. This technique reduced background noise in order to increase underwater detection accuracy. This technique avoids high computational complexity without experiencing a significant information loss (Tian et al., 2021). The average pooling layer therefore decreases the number of parameters while increasing the model’s robustness while also lowering the deviation from the computed mean. Additionally, the dropout layer is used in the two completely connected layers to prevent the overfitting issues, in which a certain number of neurons are briefly removed from the network during training. The results of the SoftMax regression are finally transferred to the (0, 1) probability interval via the output layer (Hang et al., 2019). Jian et al. (2021), reviewed multiple feature extraction and analysis techniques for underwater image processing. They also presented a benchmark underwater image database for identifying the strengths and weaknesses of the existing algorithms for underwater saliencies and images. This database

offers offer unparalleled opportunities to researchers in the underwater vision and beyond (Jian et al., 2019). Moreover, the authors in (Ullah et al., 2017; Ullah et al., 2019; Ahmad et al., 2021; Yasir et al., 2022) suggested manipulated feature extraction techniques for object detection underwater, on sea surfaces, etc. When numerous parallel convolution paths are employed, the number of social parameters is reduced. The connectivity with deepening layers can accomplish a similar (or superior) performance with fewer parameters than the connection without deepening layers. The first layer just has to focus on learning the most recent information and can learn successfully with less training data. Feature information may be separated into tiers by growing the network, improving learning efficiency.

3.1 Data acquisition

In this research, we use two different databases that are publicly available for simulations and experimental work, as depicted in Table 1. These two databases are fused together to evaluate and train the model from two different perspectives simultaneously. A few images of the underwater objects are depicted in Figure 3.

3.2 Data fusion

After selecting underwater image databases, the same dimensional data is fused using (Eq. 4). All these varying input values must be multiplied with different weights accordingly. The resultant input map can be obtained by fusing the data as depicted in (Eq. 4).

$$X_{output} = (W_{turb} \times X_{turb}) + (W_{pose} \times X_{pose}) + (W_{variety} \times X_{variety}) \quad (4)$$

Where “ \times ” is applied for element wise multiplication and X_{turb} , X_{pose} , $X_{variety}$ represent the input turbidity of water, diversity in pose, and diversity in variety, while W_{turb} , W_{pose} , $W_{variety}$ depict the learning parameters classifying numerous impact degree of these factors. The finalized value at t^{th} time

TABLE 1 Databases for underwater image processing.

| Database | Type of data | Providers |
|--------------------------|---|--|
| MUED (Jian et al., 2019) | This collection includes 8600 underwater pictures of 430 distinctive prominent items against noisy backgrounds. It is also quite different in terms of position, variety, stance, water turbidity, etc. | 1. Ocean University of China, Qingdao 2. Shandong University of Finance and Economics |
| RUIE (Liu et al., 2020) | More than 4000 underwater genuine photographs are available in this database, which is divided into three groups based on their classification: underwater image quality sub-aggregate, underwater color cast sub-aggregate, and underwater higher-level task-driven sub-aggregate. | Dalian University of Technology |

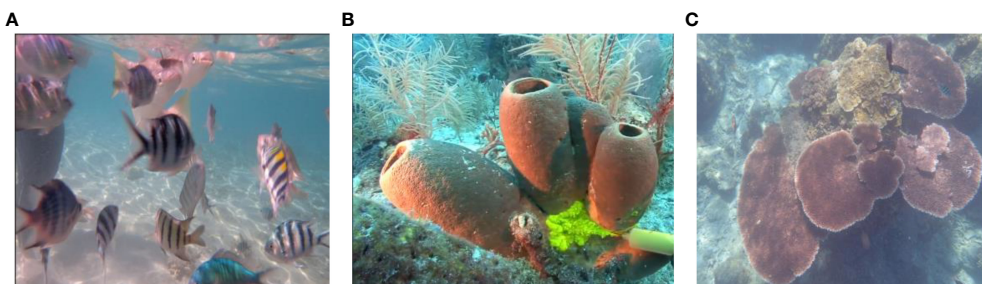


FIGURE 3
Some sample images of MUED database with a couple of subclasses.

interval is depicted by \hat{X}_t . For normalizing the output values a specific range (1, -1) is defined using the hyperbolic function depicted in (Eq. 5).

$$\hat{X}_t = \tanh(X_{output}) \quad (5)$$

This fused dataset is used for training and testing purposes. The highest accuracy rate of 95.7% is calculated for salient objects detection in underwater images.

3.3 Training and validation of the model

Different input and performance parameters used for the validation of the proposed hybrid model are depicted in Table 2.

The simulation results are calculated in Python using TensorFlow and Keras libraries for the recognition and classification of reviews. After validating using these different performance metrics an overall accuracy rate of 95.7% is achieved, which reflects the applicability of this hybrid data fusion-based model for the identification of saliencies.

4 Results and discussions

The experimental results and underlined discussion are briefly outlined in this section of the paper. To test the applicability of the proposed algorithm different performance metrics are used that are given below.

4.1 Varying training and test sets

The proposed hybrid deep learning model's recognition characteristics are evaluated using a variety of training and test sets. Figure 4 illustrates that the accuracy results grow together with the training set as can be seen. The suggested model has the greatest accuracy rate, which is 95.7%. This model's ability to recognize important things in underwater photos and video and carry out shape improvement is demonstrated by the high accuracy rate.

TABLE 2 Different parameters used for the validation and experimental work in the proposed research work.

| S. No | Parameters | Values |
|-------|-------------------------|---|
| 1. | Number of inputs | 11800 |
| 2. | Activation function | ReLu and tanh |
| 3. | Epoch size | 25 |
| 4. | Filter | 5 × 5 |
| 5. | Number of hidden layers | 5 |
| 6. | Dropout | 0.2 |
| 7. | Batch size | 32 × 32 |
| 8. | Optimizer | Adam |
| 9. | Performance metrics | ROC curve, AUC values, precision, accuracy, true-false values, varying training and test sets, miss-classification rate, simulation time. |

Training set VS time and accuracy graph

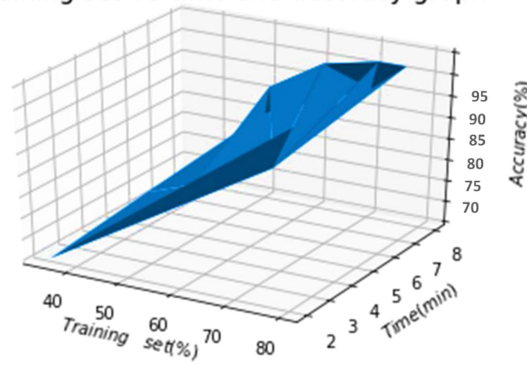


FIGURE 4 Recognition abilities based on varying training and test sets.

4.2 Varying epoch size

To evaluate the performance of the proposed hybrid deep learning system based on different epoch size. In this case an epoch size of 25 is selected for the evaluation purposes. The corresponding results of the model for the sentiment analysis are depicted in Figure 5.

From Figure 5, it is depicted that the graph contains a comparatively lower over-fitting values, which reflects the applicability of the model for the selected sentiment analysis problem.

- True Positive Rate
- False Positive Rate

True Positive Rate (TPR) is an alternate term (synonym) for recall and can be mathematically represented in (Eq. 6) as follows:

$$TPR = \frac{TP}{TP + FN} \tag{6}$$

False Positive Rate (FPR) can be mathematically represented in equation (7) as follows:

$$FPR = \frac{FP}{FP + TN} \tag{7}$$

4.3 ROC curve

A classification model’s performance is graphically depicted as a receiver operating characteristic curve (ROC curve), which is valid for all classification thresholds. This graph shows the two parameters:

At different categorization criteria, TPR vs. FPR are shown on a ROC curve. As more items are classified as positive when the classification threshold is lowered, both False Positives and True Positives rise. The hybrid model’s proposed ROC curve is shown in Figure 6.

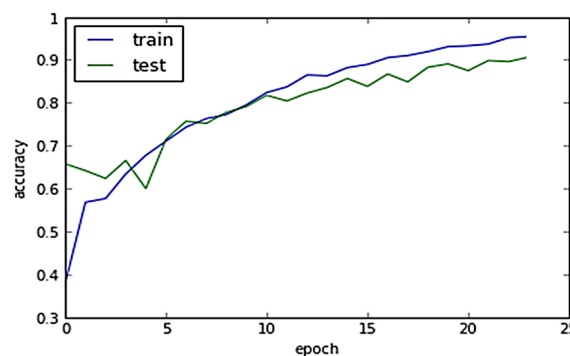
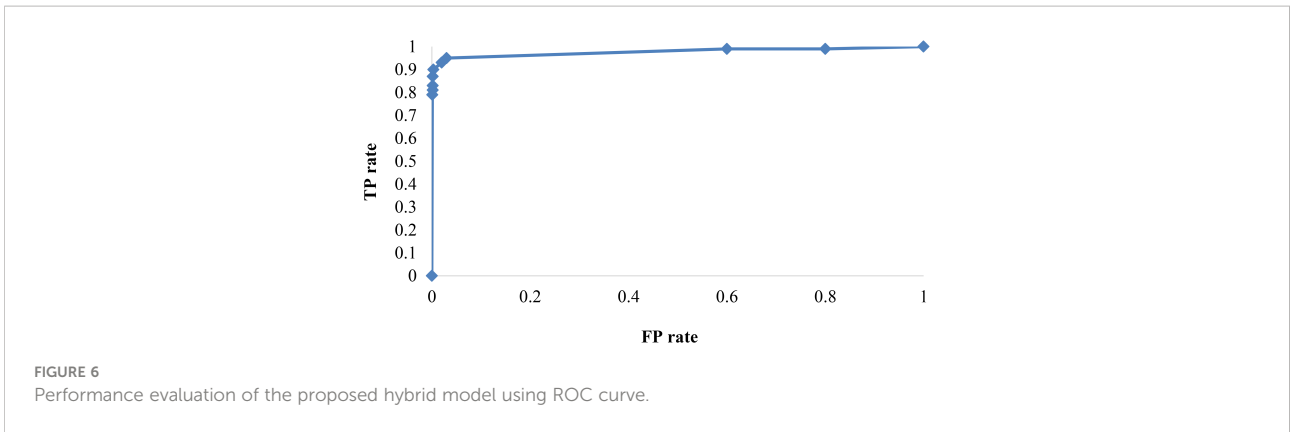


FIGURE 5 Performance evaluation of the proposed model using varying epoch size.



For validating the proposed research problem, the hybrid model is validated using different well-known baseline models reported in the proposed field. These baseline models include:

- **ARIMA** – Auto-Regressive Integrated Moving Average (ARIMA) is extensively used for time series identification problems.
- **Conv1D Net** – Significantly suggested in the extant for temporal dependent problems.
- **LSTM** (Khan and Nazir, 2022) – Long Short-Term Memory (LSTM) models are typically used for sequential data analysis and time prediction problems.
- **St-Res Net** – A sophisticated machine - learning model was developed by Zhang et al. (2017),. When compared to other algorithms, this one delivers correct outcomes for crowd and time forecasts.

The root mean square error rate (RMSE) value is selected as a validating point for comparing these benchmark algorithms

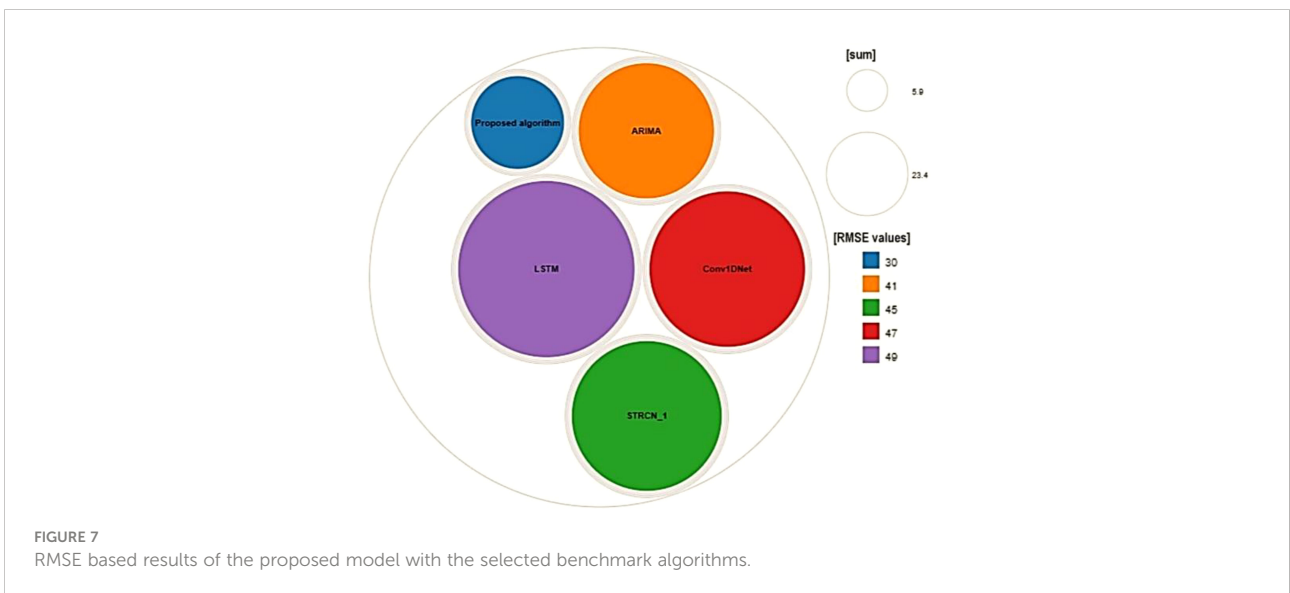
with the proposed research work. The generic followed for the RMSE value calculation is depicted in Eq. (8) as follows:

$$RMSE = \sqrt{\frac{1}{X \times T} \sum_i (x_i - \bar{x}_i)^2} \tag{8}$$

where T depicts the time interval while X represents the number of inputs observed during the corresponding iteration. x_i represents the predicted value (positive, negative, or neutral response of the user), while \bar{x}_i is the original. The RMSE-based performance results are depicted in Figure 7 below.

The proposed model’s lowest RMSE value when compared to the chosen benchmark methods demonstrates the usefulness of the presented study. The usefulness of the suggested approach for detecting prominent objects in background or underwater photos may also be deduced from this RMSE value.

For comparing the performance of the proposed hybrid model with the selected benchmark models. Accuracy is used as the performance metric in this validation process, which uses

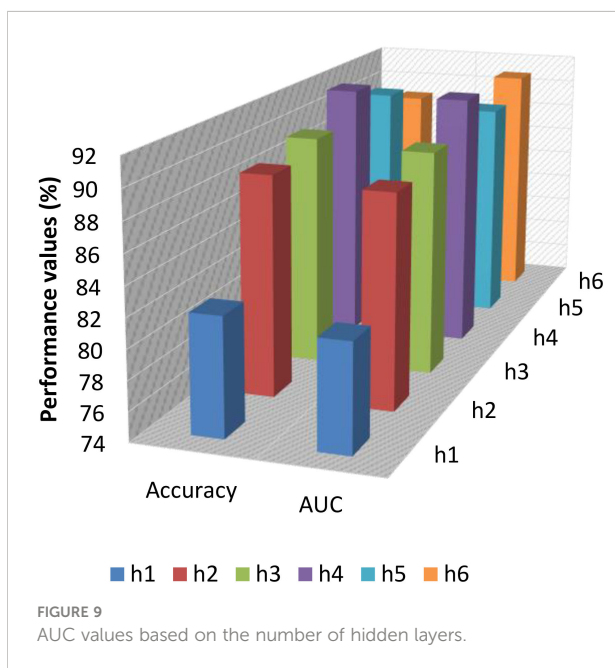
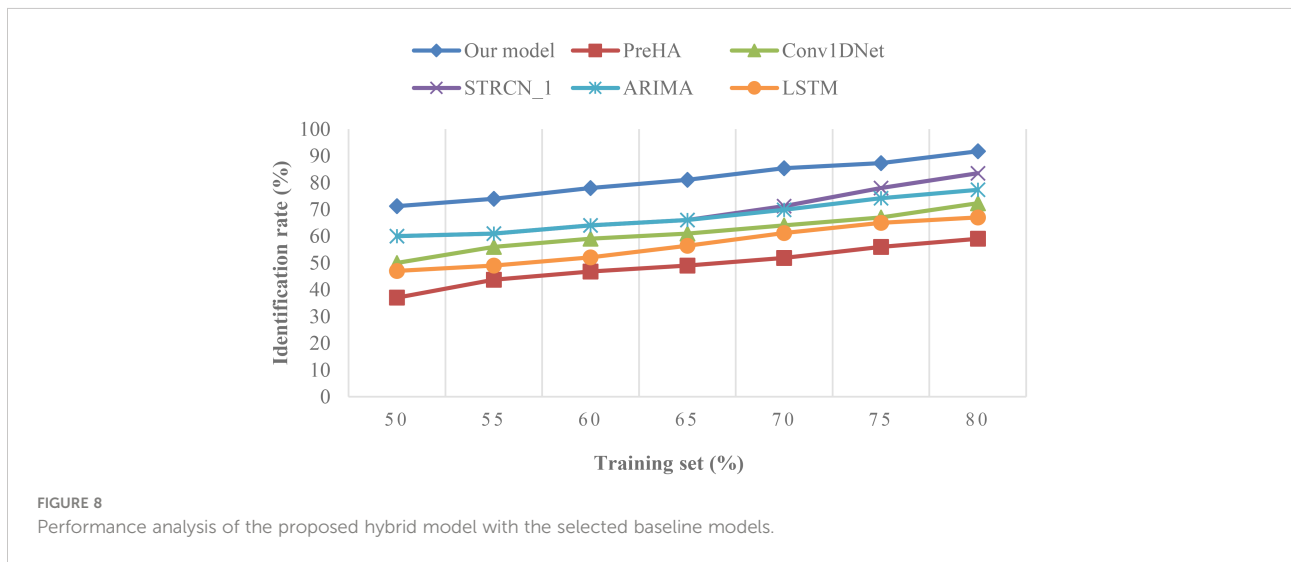


varying training and test sets. The identification capabilities of different users' emotions are depicted in Figure 8. From Figure 8 it is observed that our model outperformed by producing high identification rates based on varying training and test sets.

4.5 AUC values

The performance of the proposed model is also validated by using the area under the curve (AUC) values based on a fluctuating number of hidden layers. The underlined results are depicted in Figure 9 below.

From Figure 8, one can easily observe that the performance of the proposed model decreases after hidden layer 4 (the maximum AUC value of 95.7 is recorded at h4). This is because of the complexity of the model and the entire high simulation time. So, in our case, a deep learning structure with four hidden layers is an optimum design in our case. Figure 10 shows how much time the suggested hybrid model requires ultimately based on the various hidden layer counts. Figure 10 shows that as the number of hidden layers rises, so does the simulation duration. The circuit complexity and high simulation cost are reflected in the sudden rise in simulation time beyond hidden layer 4.



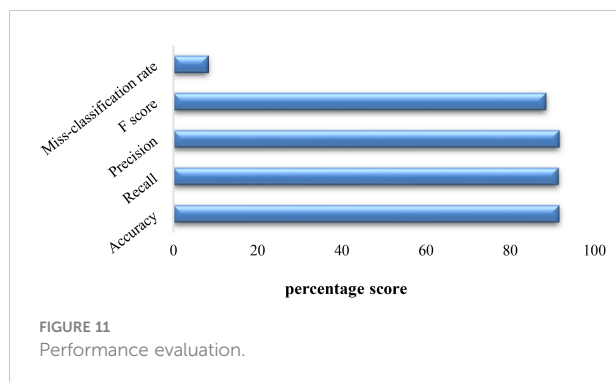
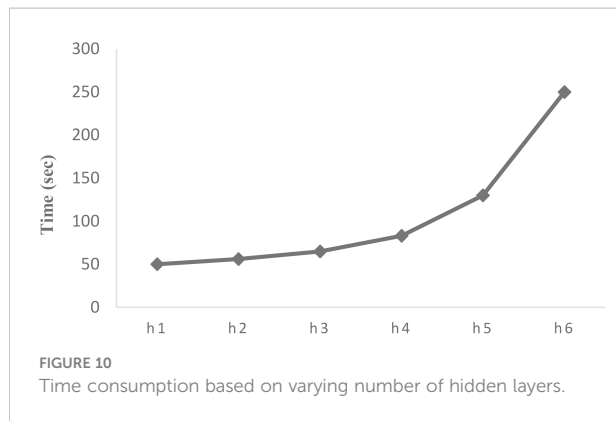
4.6 F-score, precision, and error-rate

The pertinency of the proposed model is also validated using the f-score, precision, and error rate generated. The underlined results are depicted in Figure 11 below.

The small miss-classification rate and high accuracy, precision, recall, and F-score values represented in Figure 11 reflects the pertinency of the proposed hybrid model for the optimum identification of saliencies in underwater images.

5 Conclusion

In this research, we present a hybrid deep learning model for salient object recognition in underwater photos that combines improved AlexNet and Inception v-4. The suggested model can precisely find saliencies with various sizes, water turbidities, and ocean habitats and successfully reduce the crowded backdrops thanks to the data fusion and hybrid architectural design. In order to combine data and conduct experiments, this study combines the MUED and RUIE databases. Model simulation



results are produced utilizing accuracy, computing usage, and root mean square error (RMSE). The suggested model's suitability for shape optimization and salient item recognition in underwater photos is demonstrated by its high accuracy of 95.7% and low root-mean-square errors (RMSE) values when compared to existing baseline models. The suggested hybrid model's usefulness in the intended study topic is shown by experimental findings. Our model considerably outperformed other benchmark models by producing effective outcomes with minimal training data.

References

- Ahmad, I., Rahman, T., Zeb, A., Khan, I., Ullah, I., Hamam, H., et al. (2021). Analysis of security attacks and taxonomy in underwater wireless sensor networks. *Wireless Commun. Mobile Computing*. doi: 10.1155/2021/1444024
- Cao, X., Zhang, C., Fu, H., Guo, X., and Tian, Q. (2016). Saliency-aware nonparametric foreground annotation based on weakly labeled data. *IEEE Trans. Neural Networks Learn. Syst.* 27, 1253–1265. doi: 10.1109/TNNLS.2015.2488637
- Chen, Z., Zhang, Z., Dai, F., Bu, Y., and Wang, H. (2017). Monocular vision-based underwater object detection. *Sensors* 17, 1784. doi: 10.3390/s17081784
- Cong, R., Lei, J., Fu, H., Cheng, M. M., Lin, W., and Huang, Q. (2019). Review of visual saliency detection with comprehensive information. *IEEE Trans. Circuits Syst. Video Technol.* 29, 2941–2959. doi: 10.1109/TCSVT.2018.2870832
- Gu, K., Wang, S., Yang, H., Lin, W., Zhai, G., Yang, X., et al. (2016). Saliency-guided quality assessment of screen content images. *IEEE Trans. Multimedia* 18, 1098–1110. doi: 10.1109/TMM.2016.2547343
- Halimi, A., Maccarone, A., McCarthy, A., McLaughlin, S., and Buller, G. S. (2017). Object depth profile and reflectivity restoration from sparse single-photon data acquired in underwater environments. *IEEE Trans. Comput. Imaging* 3, 472–484. doi: 10.1109/TCI.2017.2669867

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

Conceptualization, SK; methodology, SK, and IU; software, SK, and FA; validation, SK, IU, and FA; formal analysis, IU, and MS; investigation, SK; resources, IU, and FA; data curation, MS, YG, and TK; writing—original draft preparation, SK; and IU; writing—review and editing, IU, and FA; visualization, IU, YG, and TK; supervision, IU and TK; project administration, IU, and TK; funding acquisition, TK. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2020R1I1A3068305).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Hang, J., Zhang, D., Chen, P., Zhang, J., and Wang, B. (2019). Classification of plant leaf diseases based on improved convolutional neural network. *Sensors* 19, 4161. doi: 10.3390/s19194161
- Jian, M., Liu, X., Luo, H., Lu, X., Yu, H., and Dong, J. (2021). Underwater image processing and analysis: A review. *Signal Processing: Image Communication* 91, 116088. doi: 10.1016/j.image.2020.116088
- Jian, M., Qi, Q., Yu, H., Dong, J., Cui, C., Nie, X., et al. (2019). The extended marine underwater environment database and baseline evaluations. *Appl. Soft Computing* 80, 425–437. doi: 10.1016/j.asoc.2019.04.025
- Jian, M., Wang, J., Yu, H., and Wang, G. G. (2021). Integrating object proposal with attention networks for video saliency detection. *Inf. Sci.* 576, 819–830. doi: 10.1016/j.ins.2021.08.069
- Jian, M., Wang, J., Yu, H., Wang, G., Meng, X., Yang, L., et al. (2021). Visual saliency detection by integrating spatial position prior of object with background cues. *Expert Syst. Appl.* 168, 114219. doi: 10.1016/j.eswa.2020.114219
- Khan, S., and Nazir, S. (2022). Deep learning based pashto characters recognition: LSTM-based handwritten pashto characters recognition system. *Proc. Pakistan Acad. Sciences: A. Phys. Comput. Sci.* 58, 49–58. doi: 10.53560/PPASA(58-3)743
- Khan, S., Nazir, S., García-Magariño, I., and Hussain, A. (2021). Deep learning-based urban big data fusion in smart cities: Towards traffic monitoring and flow-preserving fusion. *Comput. Electrical Eng.* 89, 106906. doi: 10.1016/j.compeleceng.2020.106906
- Khan, W., Wang, H., Anwar, M. S., Ayaz, M., Ahmad, S., and Ullah, I. (2019). A multi-layer cluster based energy efficient routing scheme for UWSNs. *IEEE Access*. 7, 77398–77410. doi: 10.1109/ACCESS.2019.2922060
- Lee, G., Tai, Y. W., and Kim, J. (2018). ELD-net: An efficient deep learning architecture for accurate saliency detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 40, 1599–1610. doi: 10.1109/TPAMI.2017.2737631
- Li, C., Cong, R., Hou, J., Zhang, S., Qian, Y., and Kwong, S. (2019). Nested network with two-stream pyramid for salient object detection in optical remote sensing images. *IEEE Trans. Geosci. Remote Sens.* 57, 9156–9166. doi: 10.1109/TGRS.2019.2925070
- Liu, R., Fan, X., Zhu, M., Hou, M., and Luo, Z. (2020). Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light. *IEEE Trans. Circuits Syst. Video Technol.* 30, 4861–4875. doi: 10.1109/TCSVT.2019.2963772
- Maragos, P., Gros, P., Katsamanis, A., and Papandreou, G. (2008). “Cross-modal integration for performance improving in multimedia: A review,” in *Multimodal processing and interaction* (Springer), 1–46.
- Moghimi, M. K., and Mohanna, F. (2021). Real-time underwater image enhancement: a systematic review. *J. Real-Time Image Process.* 18, 1509–1525. doi: 10.1007/s11554-020-01052-0
- Qiao, J., Wang, X., Chen, J., and Jian, M. (2022). Low-light image enhancement with an anti-attention block-based generative adversarial network. *Electronics* 11, 1627. doi: 10.3390/electronics11101627
- Shen, J., Fan, T., Tang, M., Zhang, Q., Sun, Z., and Huang, F. (2014). A biological hierarchical model based underwater moving object detection. *Comput. Math. Methods Med.* 2014, 609801. doi: 10.1155/2014/609801
- Su, X., Ullah, I., Liu, X., and Choi, D. (2020). A review of underwater localization techniques, algorithms, and challenges. *J. Sensors*. doi: 10.1155/2020/6403161
- Tian, Y., Li, E., Liang, Z., Tan, M., and He, X. (2021). Diagnosis of typical apple diseases: A deep learning method based on multi-scale dense classification network. *Front. Plant Sci.*, 2122. doi: 10.3389/fpls.2021.698474
- Ullah, I., Chen, J., Su, X., Esposito, C., and Choi, C. (2019). Localization and detection of targets in underwater wireless sensor using distance and angle based algorithms. *IEEE Access*. 7, 45693–45704. doi: 10.1109/ACCESS.2019.2909133
- Ullah, I., Liu, Y., Su, X., and Kim, P. (2019). Efficient and accurate target localization in underwater environment. *IEEE Access*. 7, 101415–101426. doi: 10.1109/ACCESS.2019.2930735
- Ullah, I., Ming-Sheng, G. A., Kamal, M. M., and Khan, Z. (2017). “A survey on underwater localization, localization techniques and its algorithms,” in *3rd Annual International Conference on Electronics, Electrical Engineering and Information Science (EEEIS 2017)*. 252–259 (Guangzhou, Guangdong, China: Atlantis Press).
- Wang, W., Shen, J., Yang, R., and Porikli, F. (2018). Saliency-aware video object segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 40, 20–33. doi: 10.1109/TPAMI.2017.2662005
- Xu, J., Bi, P., Du, X., Li, J., and Chen, D. (2019). Generalized robust PCA: A new distance metric method for underwater target recognition. *IEEE Access* 7, 51952–51964. doi: 10.1109/ACCESS.2019.2911132
- Yasir, M., Hossain, M. S., Nazir, S., Khan, S., and Thapa, R. (2022). Object identification using manipulated edge detection techniques. *Science* 3 (1), 1–6. doi: 10.11648/j.scidev.20220301.11
- Zhang, D., Li, L., Zhu, Z., Jin, S., Gao, W., and Li, C. (2019). Object detection algorithm based on deformable convolutional networks for underwater images. *2019 2nd China Symposium Cogn. Computing Hybrid Intell. (CCHI)*, 274–279. doi: 10.1109/CCHI.2019.8901912
- Zhang, J., Zheng, Y., and Qi, D. (2017). Deep spatio-temporal residual networks for citywide crowd flows prediction. *Thirty-first AAAI Conf. Artif. Intell.* 31 (1), 1655–1661. doi: 10.1609/aaai.v31i1.10735
- Zheng, Y. (2015). Methodologies for cross-domain data fusion: An overview. *IEEE Trans. big Data* 1, 16–34. doi: 10.1109/TBDATA.2015.2465959
- Zhu, L., Guo, F., Polak, J. W., and Krishnan, R. (2018). Urban link travel time estimation using traffic states-based data fusion. *IET Intelligent Transport Syst.* 12, 651–663. doi: 10.1049/iet-its.2017.0116