JCTA Journal of Computing Theories and Applications ISSN:3024-9104

Research Article

DS-TSNet: Depth-Aware Spectral Compensation and Turbidity-Weighted Structural Stabilization with Hybrid DenseNet and Binary WGOA for Robust Underwater Image Classification

Maddukuri Srinadh * and J. B. Seventline

Department of Electrical, Electronics, and Communication Engineering, GITAM (Deemed to be University), Visakhapatnam 530045, Andhra Pradesh, India; e-mail: smadduku@gitam.in; sjoseph@gitam.edu

* Corresponding Author : Maddukuri Srinadh

Abstract: Underwater image classification remains challenging because wavelength-dependent radiance attenuation and turbidity-induced boundary distortion degrade visual features and compromise classification reliability. To address these challenges, this study proposes the Depth-Aware Spectral-Structural Learning Network (DS-TSNet), an end-to-end framework that integrates Depth-Aware Spectral Attenuation Compensation (DSACM), Turbidity-Weighted Structural Stabilization (TWSSL), hybrid feature learning using DenseNet201 and DenseNet121, feature selection through the Binary Weighted Grasshopper Optimization Algorithm (B-WGOA), and label-smoothed, class-weighted multilayer perceptron (MLP) classification. Specifically, DSACM compensates for spectral attenuation effects, whereas TWSSL enhances structural information by mitigating turbidity-induced boundary distortions. The hybrid DenseNet architecture extracts complementary features, which are subsequently refined through B-WGOA-based feature selection. This binary optimization process identifies informative feature subsets by minimizing a hybrid objective that incorporates four criteria: classification quality, stability, sparsity, and redundancy suppression. Finally, label smoothing and adaptive class weighting are incorporated into the MLP classifier to mitigate overfitting and improve probability calibration. Experimental results demonstrate that DS-TSNet achieves an accuracy of 94.71% on the Underwater Images benchmark, with a macro-F1 score of 0.9471, a balanced accuracy of 0.9471, and a Cohen's kappa coefficient of 0.9295. In a separate evaluation on the Marine_PULSE SONAR dataset, the framework achieves an accuracy of 97.89% and a Cohen's kappa coefficient of 0.9679. Furthermore, comparative evaluations against EfficientNet-B3, ResNet-50, ResNet-101, and optimization-based methods reported in the literature provide additional evidence of the framework's classification performance. Overall, these findings highlight the potential of integrating spectral compensation, structural stabilization, and optimization-guided feature selection for underwater image classification.

Keywords: Binary Weighted Grasshopper Optimization Algorithm; DenseNet201; Depth-Aware Spectral Attenuation Compensation; Feature Selection; Hybrid Deep Learning; Turbidity-Weighted Structural Stabilization; Underwater Image Classification; Underwater Image Enhancement.

Received: September, 3rd 2026
Revised: October, 29th 2026
Accepted: October, 30th 2026
Published: November, 21st 2026

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) licenses (https://creativecommons.org/licenses/by/4.0/)

1. Introduction

The classification of underwater and sonar images has become an important research topic owing to its diverse real-world applications in marine biology, environmental monitoring, defense, and autonomous navigation. Unlike conventional image acquisition, underwater imaging is subject to modality-specific degradation that affects the quality and discriminative characteristics of the acquired data. Optical underwater images commonly suffer from light scattering and wavelength-dependent absorption, resulting in reduced visibility, color distortion, and loss of fine details. In contrast, sonar images are affected by speckle noise, seabed

clutter, and ambiguous acoustic echoes, which complicate target recognition and classification [1]–[4]. These distinct degradation characteristics pose considerable challenges to the development of reliable underwater classification systems.

Such impairments affect both image enhancement and subsequent classification performance, thereby limiting the reliability of autonomous underwater applications. Traditional approaches have generally relied on handcrafted feature extraction and heuristic-based techniques for object detection and classification [5]. More recently, deep learning methods, particularly convolutional neural networks (CNNs), have demonstrated their effectiveness in learning discriminative representations from complex visual and sonar data. Nevertheless, their direct application remains challenging due to substantial intra-class variations [6], limited annotated training samples, and modality-specific image degradation [7]. These limitations highlight the need for classification frameworks that can address degraded input characteristics while maintaining informative and compact feature representations.

Recent studies have increasingly investigated preprocessing and enhancement techniques to improve image quality before classification. In optical underwater imaging, diffusion-based generative models, adversarial contrastive learning, and transformer-based U-shaped architectures [8] have demonstrated promising capabilities in recovering image details and improving color consistency. In object detection, complementary channel–spatial fusion attention (CSFA) modules have been employed to improve the detection of small objects and mitigate occlusion-related challenges [9]. Architectural developments have also explored cardinality through multi-branch feature aggregation, demonstrating that increasing the number of parallel transformations can provide performance advantages over simply increasing network depth or width at comparable computational costs [10]. In addition, recent research has combined underwater image enhancement with deep-learning-based marine-species classification, emphasizing the importance of preprocessing for automated recognition under degraded underwater imaging conditions [11]. Other studies have integrated image preprocessing, deep neural architectures, and optimization strategies for automated fish-species identification, highlighting the potential of combining feature learning with optimization-based techniques [12].

Research in sonar image analysis has followed complementary directions to address the characteristics of acoustic data. Nonparametric approaches have been investigated for sonar classification, including formulations based on convex optimization that offer interpretable modeling properties [13]. Other studies have explored robustness to distributional variations between training and test data. In particular, the attention-based complementary learning (ABCL) model [14] addresses the limited availability of active sonar datasets by combining spectrogram representations, including short-time Fourier transform and constant-Q transform features, with self-attention mechanisms and transformer architectures. Beyond classification, machine learning techniques have also been applied to forward-looking sonar (FLS) odometry to estimate ego-motion in synthetic and real-world scenarios using CNN- and recurrent convolutional neural network (RCNN)-based models [15]. These developments demonstrate the expanding role of deep learning in underwater perception, including classification, navigation, and localization tasks relevant to autonomous underwater vehicles (AUVs) and remotely operated vehicles (ROVs).

Despite these advances, several challenges remain in developing reliable underwater classification frameworks. First, image degradation can weaken discriminative features, while conventional feature extraction may produce high-dimensional representations containing redundant or less informative information. Second, although deep learning architectures and attention mechanisms improve representation learning, they do not necessarily ensure that the resulting feature representations are compact and suitable for downstream classification. Third, feature selection and optimization strategies are needed to balance classification performance with feature-subset characteristics, including sparsity, stability, and redundancy. These issues motivate the development of an integrated framework that combines task-specific image preprocessing, complementary deep feature extraction, and optimization-guided feature selection.

To address these challenges, this study proposes the Depth-Aware Spectral-Structural Learning Network (DS-TSNet), an integrated framework for underwater image classification. The framework combines task-specific spectral compensation and structural stabilization with hybrid deep feature learning and binary optimization-guided feature selection. The main contributions of this study are summarized as follows:

1. A task-specific spectral compensation module, termed Depth-Aware Spectral Attenuation Compensation (DSACM), is introduced to address color imbalance and radiometric degradation in underwater images. It combines channel-wise compensation, gray-world-inspired normalization, gamma correction, contrast enhancement, and denoising to improve the input representation for subsequent feature extraction.
2. A Turbidity-Weighted Structural Stabilization (TWSSL) module is developed to enhance local boundary and texture information affected by turbidity. The module incorporates gradient-based structural descriptions, edge detection, bilateral filtering, and detail enhancement to support the extraction of more informative visual features.
3. A binary optimization-guided feature selection procedure based on the Binary Weighted Grasshopper Optimization Algorithm (B-WGOA) is proposed to identify a discriminative subset from the hybrid feature representation. Its hybrid objective considers classification quality, stability, sparsity, and redundancy suppression, with the aim of retaining informative features while reducing unnecessary dimensionality.

The proposed framework further integrates established techniques, including DenseNet201–DenseNet121-based hybrid feature extraction, variance and mutual-information filtering, and a label-smoothed, class-weighted multilayer perceptron (MLP) classifier. These components provide the feature learning and classification backbone of DS-TSNet, whereas the principal methodological contributions lie in the task-specific spectral compensation, structural stabilization, and binary optimization-guided feature selection. The effectiveness of the integrated framework and its constituent components is investigated through comparative evaluations and component-wise ablation experiments. The remainder of this paper is organized as follows. Section 2 reviews related work and identifies the research gaps. Section 3 presents the proposed DS-TSNet framework. Section 4 describes the experimental setup and discusses the evaluation results. Section 5 discusses the limitations of the study, and Section 6 concludes the paper.

2. Related Work and Research Gaps

The range of underwater applications has expanded considerably in recent years, particularly in target recognition and sonar seabed-sediment classification. This progress has been driven largely by the adoption of deep learning algorithms and optimization techniques for underwater and sonar image analysis. Early studies primarily focused on feature fusion and evolutionary optimization. For instance, study [16] proposed a CNN-BiGRU classifier incorporating residual and attention blocks, in which sonar texture and highlight features were encoded using genetic techniques and subsequently reduced. The method achieved an accuracy of 99.1% with rapid convergence. However, its evaluation was conducted on a relatively small-scale dataset, limiting the evidence of its applicability to broader scenarios. Subsequent research has explored transformer-based architectures to improve recognition capabilities. Specifically, study [17] combined Faster R-CNN with Swin Transformer and Path Aggregation Network (PAN), employing ROI Align for underwater object detection. The method achieved an accuracy of 80.54% on the URPC2018 dataset, particularly in detecting small objects, although its computational complexity may restrict real-time deployment. In addition, hybrid architectures inspired by quantum computing have been investigated. Study [18] integrated classical backbones, including ResNet50, ResNet18, and InceptionV3, with quantum circuits. The resulting ResNet50-QCNN achieved 88% accuracy, demonstrating the potential of hybrid classical–quantum architectures for underwater animal identification and classification. Table 1 summarizes representative studies on underwater image classification, object recognition, and sonar image analysis, highlighting their respective approaches, application settings, and limitations.

Beyond the approaches summarized in Table 1, previous research has investigated several complementary strategies to address the limitations of underwater and sonar image analysis. A comprehensive review in [19], covering more than 60 studies on sonar imaging, identified the predominance of CNN-based approaches and the increasing use of GAN-based data augmentation, while highlighting the limited availability of benchmarking datasets, particularly for vision transformer-based methods. To address the scarcity of labeled data, study [20] proposed a semi-synthetic data generation framework combining optical shape extraction, Weibull distribution, and VGG19-based transfer learning. Although the method

achieved 97.76% accuracy, limitations related to synthetic sample quality and class imbalance remained. Similarly, uncertainty-driven active learning has been explored to improve sonar classification under limited annotation conditions. In [21], an attention-based complementary learning framework employed deep ensemble variance for sample selection and demonstrated promising classification performance across different ocean environments. Nevertheless, its reliance on expert annotation and its primary focus on binary target/non-target classification limit its applicability to more diverse classification tasks.

Table 1. Research-Gap Synthesis from Representative Underwater Classification and Recognition

Ref.	Representative approach	Primary task / setting	Reported limitation	Corresponding research gap
[16]	CNN–BiGRU with attention, feature fusion and gene encoding	Underwater target classification	Validation was conducted on a relatively small-scale dataset	High-dimensional feature representation and limited evaluation
[17]	Faster R-CNN with Swin Transformer and PAN	Underwater object detection on URPC2018	High computational complexity limits practical real-time deployment	Limited efficiency and heterogeneous evaluation
[18]	Hybrid classical–quantum architecture using ResNet backbones	Underwater animal identification and classification	Evaluation is tied to a specific hybrid architecture and dataset setting	Representation robustness and heterogeneous evaluation
[19]	Review of deep-learning methods for passive sonar imagery	Target recognition, detection and segmentation	Highlights limited benchmarking datasets and the continued need for broader evaluation	Limited evaluation across datasets and underwater settings
[20]	Deep transfer learning with semi-synthetic training data	Side-scan sonar object classification	Semi-synthetic samples may have quality limitations and class imbalance	Feature/representation robustness and dataset limitations
[21]	Attention-based complementary learning with uncertainty-driven active learning	Active sonar target classification	Requires expert labeling and is primarily formulated for binary target/non-target classification	Limited heterogeneous evaluation and task-specific applicability
[22]	Automated visual attention, LDA fusion and active-contour segmentation	Manufactured-object detection/segmentation in underwater colour images	Focuses primarily on object localization and segmentation rather than integrated classification	Structural degradation and task-specific processing
[23]	CNN-based modulation classification using image representations	Underwater acoustic communication	Evaluation is performed under controlled communication-channel conditions	Limited generalization across heterogeneous underwater conditions
[24]	Survey of deep-learning-based computer vision for sonar imagery	Sonar recognition, detection and segmentation	Identifies dataset, modality and benchmarking challenges across sonar vision	Limited heterogeneous evaluation
[25]	UWIE, SURF, Firefly optimization and neural-network classification	Fish-species classification	Performance is dependent on the evaluated dataset and species configuration	Feature selection, representation compactness and generalization
[26]	Deep-learning framework for small side-scan sonar sediment classification	Sonar sediment classification	Focuses on small-image datasets and task-specific classification conditions	Limited evaluation and representation robustness

Other studies have explored different feature representation and recognition strategies across underwater sensing applications. Study [22] investigated automated visual attention, linear discriminant analysis-based feature fusion, and active-contour segmentation for underwater object analysis. In underwater acoustic communication, study [23] developed a CNN-based automatic modulation classification framework using image representations of preprocessed signals, achieving 99% accuracy at signal-to-noise ratios (SNRs) above 10 dB. However, its evaluation under controlled OFDM communication-channel conditions limits the extent to which the results can be generalized to more heterogeneous underwater environments. Furthermore, study [24] investigated automated detection and segmentation of man-made underwater objects by integrating visual attention, feature fusion, and active-contour

modeling. For fish-species classification, study [25] combined underwater image enhancement (UWIE), SURF feature extraction, Firefly-based region-of-interest optimization, and neural-network classification, achieving 99% accuracy on approximately 10,000 reef-fish images across five species. Despite its reported efficiency, the approach remains dependent on the evaluated dataset and species configuration. Meanwhile, study [26] explored deep learning and data generation strategies for sonar sediment classification, reporting promising results using synthetic data, fine-tuning, and hybrid training configurations. However, the reliability and applicability of synthetic data remain important considerations, particularly in specialized sonar applications.

Overall, the reviewed studies demonstrate substantial progress in underwater recognition through feature fusion, attention mechanisms, transformer architectures, active learning, data augmentation, and optimization-based classification. However, several challenges remain relevant to the development of reliable underwater image classification frameworks. First, spectral and visibility distortions caused by underwater light absorption and scattering are not explicitly addressed by many existing recognition approaches Spectral and visibility distortions: The majority of the reviewed recognition approaches [17], [18], [20], [25]. Second, turbidity and scattering can degrade object boundaries and local texture information, yet structural stabilization is not consistently incorporated into classification-oriented preprocessing strategies [17], [18], [21], [22], [24]. Third, deep feature extraction may produce high-dimensional representations containing redundant or less informative features, highlighting the need for feature selection strategies that balance classification performance with representation compactness [16], [20], [25]. Finally, differences in sensing modalities, object categories, acquisition conditions, and dataset characteristics complicate performance comparison and evaluation across underwater recognition tasks [16], [18]–[21], [25], [26].

Based on these observations, this study proposes DS-TSNet to address the identified challenges through an integrated classification framework. DSACM is introduced to compensate for spectral and radiometric degradation, while TWSSL aims to preserve structural information affected by turbidity. Hybrid DenseNet201–DenseNet121 feature extraction, combined with variance and mutual-information filtering, provides complementary feature representations, which are further refined using B-WGOA to reduce feature dimensionality and redundancy. The resulting features are classified using a label-smoothed, class-weighted MLP. The effectiveness of the proposed framework is investigated through component-wise ablation experiments and separate evaluations on optical underwater and sonar datasets. These evaluations examine the applicability of the framework to different underwater sensing modalities but do not establish direct cross-dataset transfer.

3. The Proposed DS-TSNet Framework

This section highlights the DS-TSNet framework, a spectral-structural learning architecture that enhances the classification of underwater images in challenging visibility scenarios owing to attenuation, turbidity, and scattering phenomena in underwater environments. The approach involves physics-based spectral correction, structure stabilization, deep feature extraction, and an optimal feature selection process.

The numerical implementation settings used for image preprocessing, dataset partitioning, hybrid DenseNet feature extraction, and statistical feature pre-filtering are listed in Table 2. These numbers summarize the main characteristics of the considered experiments, including the input resolution, normalization parameters, dataset sizes, dimensionalities of DenseNet201 and DenseNet121 features, resulting hybrid representation, and successive variance–mutual-information reduction stages.

3.1. Overview of the Proposed DS-TSNet Framework

Figure 1 depicts the overall working mechanism of the proposed DS-TSNet for underwater image classification. It starts with preprocessing techniques, such as resizing and normalization, which is then followed by the application of the DSACM module to handle the wavelength-dependent color shift issue owing to illumination variations. Next is the TWSSL, which helps enhance edge details and stabilizes structural information that can be affected by scattering and backscatter noises. Subsequently, the Dense-Net201-DenseNet121 model was applied to extract multiple scales of information from the image data. Variance filter-based feature selection and mutual information-based selection help reduce the feature space to a

more compact set, which is then selected using the B-WGOA. Finally, the label-smoothed weighted MLP classifier determines the final class.

Table 2. Numerical Implementation Configuration of the DS-TSNet Framework

Parameter	Numerical value
Random seed	42
Input height	128
Input width	128
Input channels	3
ImageNet mean – R	0.485
ImageNet mean – G	0.456
ImageNet mean – B	0.406
ImageNet standard deviation – R	0.229
ImageNet standard deviation – G	0.224
ImageNet standard deviation – B	0.225
Training images	3,480
Images per training class	870
Internal validation proportion	10%
External test images	870
Echinus test images	218
Holothurian test images	218
Scallop test images	217
Starfish test images	217
DenseNet201 feature dimension	1,920
DenseNet121 feature dimension	1,024
Raw hybrid feature dimension	2,944
Variance-stage features	2,400
Mutual-information neighbours	3
Final prefiltered features	1,200

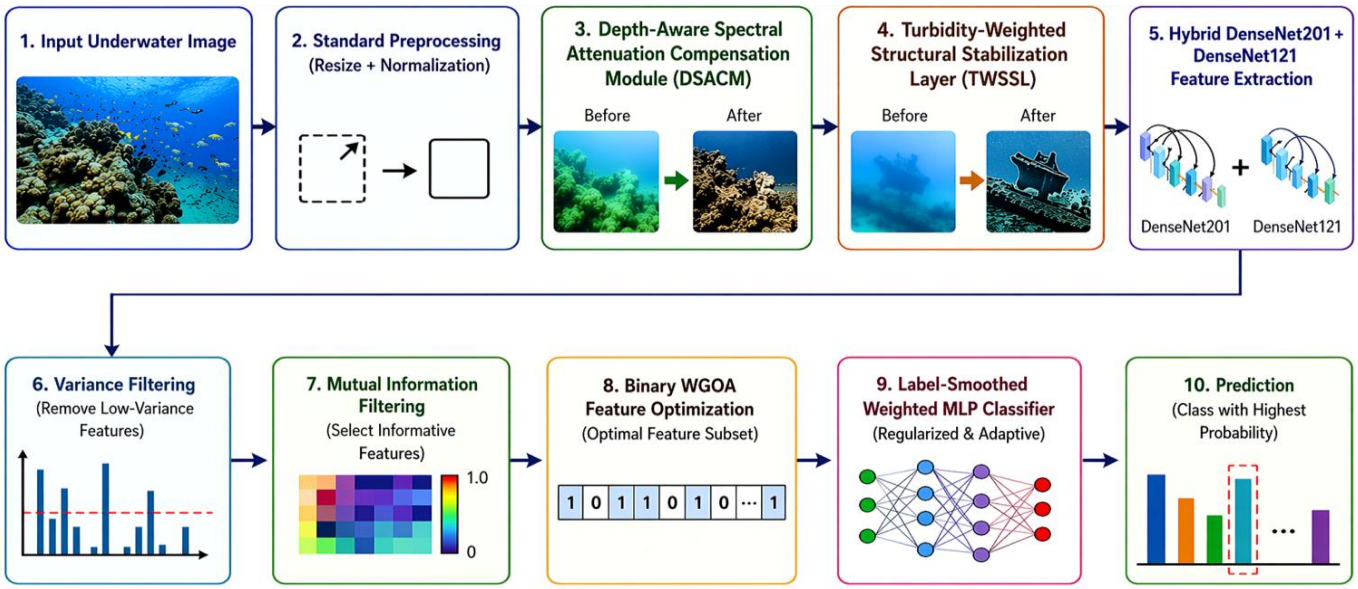


Figure 1. Overall Architecture of the proposed DS-TSNet Framework for Underwater Image Classification

3.2. Standard Preprocessing

This phase begins with the underwater image acquisition process, where RGB images denoted by $I(x, y, c)$, with $c \in \backslash$ being one of R, G, or B, are acquired at different illumination levels, turbidity levels, and depths. In this case, the signal is affected by light attenuation

and scattering effects based on the wavelengths in water. Therefore, this signal can be represented as in Eq. 1

$$I_{x,y,c} = J_{x,y,c}t_{x,y,c} + B_{x,y,c} \quad (1)$$

where $J(x, y, c)$ is the scene radiance, $t(x, y, c)$ refers to the attenuation of the light transmission, and $B(x, y, c)$ is the backscatter noise. Standard processing is used to normalize the spatial and radiometric properties before enhancing the images. The images were resampled to a particular size of $H \times W$ and normalized based on their intensities according to Eq. (2):

$$I'_{x,y,c} = \frac{I_{x,y,c} - \mu_c}{\sigma_c} \quad (2)$$

3.3. Depth-Aware Spectral Attenuation Compensation Module (DSACM)

The Depth-Aware Spectral Attenuation Compensation Module (DSACM) was designed to address chromatic imbalance and radiometric distortions in underwater images. The module applies channel-wise gain adjustment based on the general spectral attenuation characteristics of water. It should be noted that no explicit scene-depth estimation or physical parameterization was performed in this study. Instead, spectral compensation was implemented using predefined channel gains representing the relative attenuation characteristics of underwater RGB channels. The overall processing concept of DSACM, including the underwater image distortions addressed by the module and the resulting enhanced image, is illustrated in Figure 2.

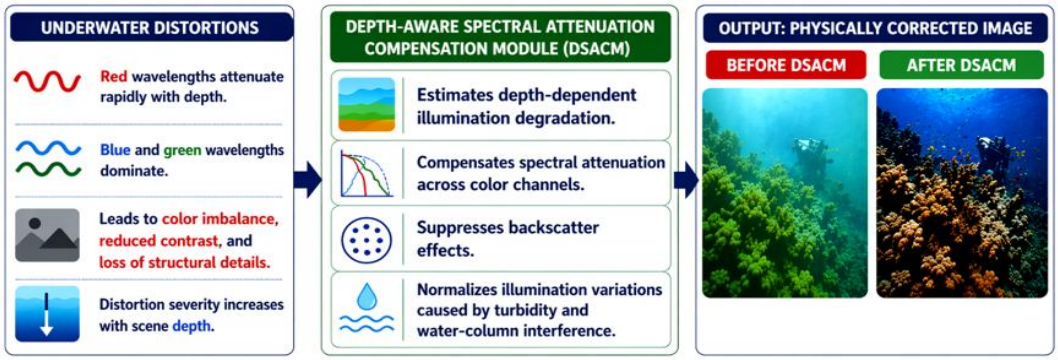


Figure 2. Overview of the DSACM and its Underwater Image Enhancement Process.

The initial channel-wise compensation and intensity normalization steps aim to reduce color imbalance while preserving the underlying object structures. As summarized in Table 3, the compensation gains were set to 1.90, 1.12, and 1.00 for the red, green, and blue channels, respectively.

Table 3. Detailed Implementation Configuration of the DSACM

Component	Implemented configuration
DSACM spectral gains	Red = 1.90, Green = 1.12, and Blue = 1.00
Channel balancing	Gray-world normalization using global channel means
Gamma correction	(γ=0.90)
CLAHE	Clip limit of 2.2 and tile grid of 8×8
Denoising	Fast Non-Local Means, with h=5, $h_{h_{\text{color}}}$=5, template window = 7 × 7, and search window = 21 × 21
Depth map	Not explicitly measured/estimated in implementation
Physical attenuation coefficient (b)	Not independently calibrated

The higher gain applied to the red channel compensates for its relatively greater attenuation in underwater environments, while the blue channel serves as the reference channel.

The subsequent processing stages include gamma correction with γ=0.90, contrast enhancement using contrast-limited adaptive histogram equalization (CLAHE) in the LAB color space with a clip limit of 2.2 and an 8×8 tile grid, and noise reduction using Fast Non-Local Means with h=5 and h_{color}=5, a 7×7 template window, and a 21×21 search window. Finally, the processed image intensities are clipped to the [0,1] range before being forwarded to TWSSL for further structural stabilization. The spectral compensation operation is defined in Eq. (3):

$$I'_c(x,y) = \text{clip}(g_c I_c(x,y), 0,1) \tag{3}$$

where $I_c(x,y)$ denotes the normalized intensity of channel c, and g_c represents the predefined gain factor applied to that channel. These gains were selected as fixed enhancement factors based on the general relative attenuation characteristics of underwater RGB channels, with greater compensation applied to the red channel and the blue channel retained as the reference.

The application of DSACM produces spectrally adjusted underwater image representations for subsequent processing and feature extraction. By combining predefined channel-wise compensation gains, gray-world normalization, gamma correction, local contrast enhancement, and denoising, the module aims to mitigate wavelength-dependent color imbalance and radiometric distortions in the input images. These operations do not require explicit scene-depth estimation or independently calibrated physical attenuation coefficients. Furthermore, the combined enhancement operations are intended to reduce image noise and preserve relevant boundary and texture details, providing more consistent input representations for the subsequent structural enhancement stage. Consequently, the DSACM-processed images are used as inputs to the hybrid DenseNet feature extraction stage, with the aim of reducing color distortions and unwanted intensity variations that may affect the learning of discriminative semantic and structural features.

3.4. Turbidity-Weighted Structural Stabilization Layer (TWSSL)

The proposed Turbidity-Weighted Structural Stabilization Layer (TWSSL) is designed to mitigate scattering-related degradation, boundary blurring, and texture distortion in underwater images. As illustrated in Figure 3, TWSSL consists of seven processing stages: turbidity-influence estimation, adaptive structural weighting, structural reinforcement and noise suppression, structural stabilization, edge and texture enhancement, regional contrast balancing, and stabilized image generation. These stages are designed to preserve local structural information and enhance boundary details that may be degraded by underwater imaging conditions.

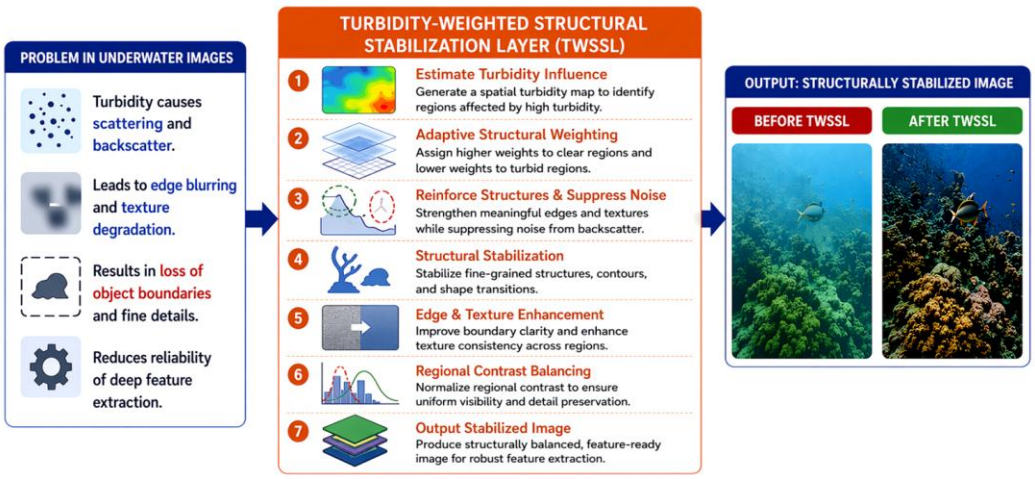


Figure 3. TWSSL for Edge-Preserving Structural Enhancement in Underwater Images

First, the DSACM-enhanced image is converted into grayscale, and its local structural information is extracted using the Sobel operator. The horizontal and vertical gradient responses are then used to calculate the gradient magnitude at each pixel. A spatial turbidity-

influence proxy is subsequently estimated from the inverse normalized gradient magnitude and smoothed using a two-dimensional Gaussian filter, as expressed in Eq. (4):

$$T(x,y) = G_\sigma * \text{clip}\left(1 - \frac{\sqrt{G_x(x,y)^2 + G_y(x,y)^2}}{G_{\max} + \varepsilon}, 0,1\right) \quad (4)$$

where $G_x(x,y)$ and $G_y(x,y)$ denote the horizontal and vertical gradient responses at pixel (x,y), $G_{\max}$ represents the maximum gradient magnitude in the image, while ϵ is a small positive constant introduced to prevent division by zero. The operator $\text{clip}(\cdot,0,1)$, constrains the normalized values to the range [0,1], and G_σ denotes a two-dimensional Gaussian smoothing kernel with standard deviation σ. The asterisk (*) represents the convolution operation..

In the second stage, adaptive structural weighting is performed with the estimated turbidity influence using Eq.5:

$$W(x,y) = \frac{1}{1 + \gamma T(x,y)} \quad (5)$$

where g=1.5 determines the sensitivity of the structure. This function provides higher weights for relatively clear structures and lower weights for regions with significant responses of the turbidity proxy.

The third and fourth steps involve structural reinforcement, noise suppression, and stabilization. Canny edge detection with thresholds 50 and 150, followed by 3×3 Gaussian smoothing, identifies the prominent features of the objects. Bilateral filtering d =7, σ_{color} =σ_{space}=60 suppresses noise while preserving edge transitions, whereas unsharp masking (strength 1.15, Gaussian σ=2.0) and edge reinforcement (weight 1.90) improve boundary continuity. The filtered and reinforced representations are adaptively combined and represented using Eq. (6):

$$I_f(x,y) = W(x,y)I_{\text{filtered}}(x,y) + [1 - W(x,y)]I_{\text{reinforced}}(x,y) \quad (6)$$

where I_{filtered} denotes the bilateral-filtered representation, and $I_{\text{reinforced}}$ represents the structurally enhanced image. This fusion retains more filtered information in regions with stronger structural clarity and increases reinforcement in regions with higher turbidity-proxy responses.

In the fifth stage, edge and texture enhancement is conducted to enhance the fine-grained structural information. The local details are extracted using Gaussian smoothing and then enhanced by Eq. (7):

$$I_{\text{detail}}(x,y) = \text{clip}\left[I_f(x,y) + \lambda_d\left(I_f(x,y) - G_{\sigma=1.2} * I_f(x,y)\right), 0,1\right] \quad (7)$$

where λ_d=1.25 controls local texture enhancement. Next, regional contrast balancing is conducted using the LAB-space CLAHE with a clip limit of 2.0 and an 8×8 tile grid to improve the local visibility and intensity consistency. Finally, the structurally stabilized image is produced and sent to the hybrid DenseNet201-DenseNet121 feature extraction stage.

Table 4. Detailed Implementation Configuration of the TWSSL

Component	Implemented configuration
TWSSL turbidity proxy	$T(x,y) = 1 - \frac{G(x,y)}{\max(G)}$, where $G(x,y)$ is the normalized Sobel-gradient magnitude.
Turbidity smoothing	Gaussian blur, 5×5 kernel
Edge detector	Canny edge detection with thresholds 50/150
Edge smoothing	Gaussian blur, 3×3 kernel
Structural filter	Bilateral filter
Bilateral parameters	Diameter d=7, color sigma σ_{color}=60, and spatial sigma σ_{space}=60
Edge weighting	1.90
Detail weighting	1.25
Sharpening	Unsharp masking with strength = 1.15 and Gaussian standard deviation σ=2.0
Final contrast normalization	CLAHE with clip limit 2.0 and tile grid 8×8

The implementation parameters and configurations used throughout the TWSSL processing stages are summarized in Table 4. These settings specify the turbidity-proxy estimation, filtering operations, edge enhancement, detail reinforcement, and final contrast normalization. The overall TWSSL process combines image-derived turbidity-influence estimation, adaptive structural weighting, edge-preserving filtering, boundary reinforcement, and regional contrast enhancement. Through these operations, TWSSL produces enhanced image representations intended to preserve local structural information and improve the quality of subsequent hybrid feature extraction.

3.5. Hybrid DenseNet201–DenseNet121 Feature Extraction

The stabilized images are processed through a hybrid DenseNet201–DenseNet121 backbone to extract hierarchical semantic representations. Dense connectivity enables feature reuse across layers according to Eq. 8

$$\mathbf{x}_l = H_l([\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{l-1}]) \tag{8}$$

where $H_l(\cdot)$ represents composite convolutional transformation and bracket notation indicates feature concatenation. Multi-scale embeddings obtained from both architectures are fused as shown in Eq. 9

$$\mathbf{F} = P\mathbf{F}_{201} + (1 - P)\mathbf{F}_{121} \tag{9}$$

where P controls fusion balance. This integration captures both global semantic ab-straction and fine-grained structural texture information.

A hybrid deep feature extraction and optimization model that utilizes the DenseNet201 and DenseNet121 architectures was implemented. In this step, a 2,944-dimensional (1,920 + 1,024) feature vector was designed by concatenating the outputs of DenseNet201 and DenseNet121.

3.6. Two-Stage Prefiltering Strategy

The proposed two-stage filtering method integrates variance-based filtering with mutual information-based filtering to achieve an optimal and highly discriminative subset of features before optimization. The first stage employs variance-based filtering to remove low-variance features with low statistical variation across instances, as such features offer minimal separability among different underwater objects and may add noise or redundancy to the learning process. It retains 2400 features. This approach guarantees that features that are sensitive to intensity variation and structural response are considered in further processing steps. This feature variance was computed as in Eq. 10

$$\sigma_j^2 = \frac{1}{N}\sum_{i=1}^{N}\left(x_{ij} - \bar{x}_j\right)^2 \tag{10}$$

The top features with largest variance were retained. From this reduced set, mutual information filtering evaluates the dependency of the selected features with the corresponding classes and reduces the features to 1200, where maximum relevance is ensured with minimal redundancy among the selected features. By enforcing both dispersion sensitivity and class-dependency consideration, this two-stage filtering technique provides an optimal intermediate feature representation for optimizing the performance of the B-WGOA in achieving the optimal feature subset.

3.7. Binary Weighted Grasshopper Optimization Algorithm (Binary WGOA)

The B-WGOA-based feature-selection procedure follows the sequential workflow shown in Figure 4. Step 1: Prefiltering reduces the original feature space to the top-K candidate features, with K=1200. Step 2: Fitness Evaluation evaluates each candidate subset using SVM-based cross-validation. Step 3: Composite Score Determination calculates the fitness value using the composite objective. Step 4: B-WGOA Initialization generates the initial population of binary candidate solutions, where each binary position represents the selection or rejection of an individual feature.

Step 5: The B-WGOA optimization loop iteratively searches for the feature subset with the lowest fitness. Step 5.1 initializes the optimization at t=1, and proceeds up to the maximum iteration T. Step 5.2 updates candidate positions according to the exploration and exploitation rules to balance the global search and local refinement. Step 5.3 applies Lévy-flight perturbation to introduce occasional long-range movements and improve escape from local optima. The Lévy perturbation is applied with probability 0.10, using the Mantegna Lévy-flight formulation with beta=1.5 and a perturbation scaling factor of 0.01. Step 5.4 applies binary mutation to introduce small random changes in the feature-selection states, with a mutation probability that decreases from 0.15 to 0.02 during optimization. Step 5.5 evaluates the resulting binary mask using the composite fitness function. Step 5.6 compares the current solution with the global best solution and updates the best feature mask whenever a lower fitness value is obtained. Step 5.7 checks the stopping condition; if the criterion is not satisfied, the optimization proceeds to the next iteration, whereas satisfaction of the criterion terminates the search. Step 6: Best Feature Mask Selection retains the binary mask corresponding to the minimum fitness value obtained throughout the optimization process. Step 7: The Final Feature Subset Output maps the selected binary positions to their corresponding feature indices and generates the final reduced feature representation for subsequent classification.

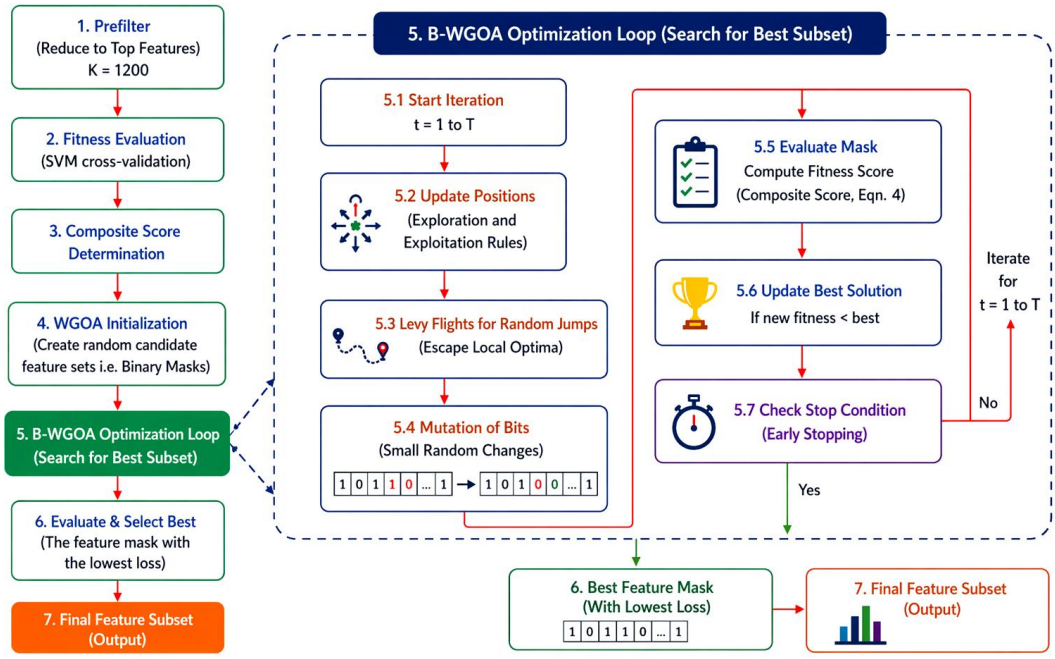


Figure 4. Integrated Workflow of the Proposed B-WGOA–Based Feature Optimization Framework for Optimal Feature Subset Selection

To enhance global search capability and reduce the risk of premature convergence to local optima, Lévy-flight-based random perturbations are applied, followed by bit mutations to maintain population diversity. Each candidate feature subset is represented by a binary mask $m \in \{0,1\}^{D_{\mathrm{pf}}}$, where the selected features are denoted by $\mathbf{X}_{\mathrm{pf}}^{(m)}$. Each mask is evaluated using repeated stratified cross-validation with an RBF-SVM classifier. The evaluation score combines multiple classification performance metrics, as defined in Eq. (11).

$$\mathrm{Score}(m) = \alpha\,\overline{\mathrm{Acc}}(m) + \beta\,\overline{F1}_{\mathrm{macro}}(m) + \gamma\,\overline{\mathrm{BalAcc}}(m) - \delta\,\sigma_{\mathrm{CV}}(m) \tag{11}$$

where $\overline{\mathrm{Acc}}(m)$, $\overline{F1}_{\mathrm{macro}}(m)$, and $\overline{\mathrm{BalAcc}}(m)$ denote the average accuracy, macro-F1 score, and balanced accuracy across the cross-validation folds, respectively. The term $\sigma_{\mathrm{CV}}(m)$ represents the variance in classification performance across the cross-validation folds and accounts for performance variability. In addition, two structural penalties are incorporated into the optimization objective: a sparsity term proportional to the ratio of selected

features to the total number of prefiltered features, $\|m\|_0/D_{\text{pf}}$, and a redundancy term represented by the average absolute correlation among the selected features, denoted by $\rho\left(\mathbf{X}_{\text{pf}}^{(m)}\right)$. The final objective minimized by the optimizer is defined in Eq. (12).

$$L_{m,t} = 1 - \text{Score}(m) + \lambda_t \frac{\|m\|_0}{D_{\text{pf}}} + \eta\, \rho\left(\mathbf{X}_{\text{pf}}^{(m)}\right) \tag{12}$$

where $L_{m,t}$ denotes the objective value of the candidate mask m at iteration t, λ_t represents the iteration-dependent sparsity penalty coefficient, and η controls the strength of the redundancy penalty. The sparsity coefficient λ_t increases linearly with iteration t, encouraging the optimizer to progressively favor compact feature subsets while maintaining classification performance.

The RBF-SVM used within the B-WGOA is an optimization surrogate for evaluating feature subsets, rather than the final prediction model. During optimization, each candidate binary feature mask is evaluated using the selected features, standardized inputs, and repeated stratified cross-validation to enable the search process to estimate the discriminative quality of alternative feature subsets in a computationally stable manner. After the optimal feature subset is identified, the selected representation is passed to the final LSW-MLP classifier, designed to learn nonlinear class boundaries incorporating class weighting and label smoothing to enable better class-balanced learning and reduce prediction overconfidence. As such, the SVM and LSW-MLP have different methodological roles: the former provides a consistent and comparatively lightweight criterion for guiding the stochastic feature-selection search, while the latter performs the final discriminative classification. This also decouples the feature-selection process from being intrinsically tied to the architecture of the final classifier and enables the selected feature representation to be evaluated by a complementary nonlinear learning model.

3.8. Label-Smoothed Weighted MLP–Based Classification and Final Prediction

The final classification step uses a Label-Smoothed Weighted Multilayer Perceptron (LSW-MLP) to generate predictions from the feature subset selected by B-WGOA. Unlike conventional MLP classifiers that assign equal importance to all classes, the proposed approach uses class-weighted learning to address class imbalance and incorporates label smoothing to reduce prediction overconfidence. The feature set is passed through a fully connected hidden layer and an output layer. The softmax function then produces posterior class probabilities, and the class with the highest probability is selected.

The final classification step makes use of an MLP type, working on the feature subset discovered by the B-WGOA algorithm. The MLP comprises one hidden fully connected layer with 512 units, Batch Normalization, ReLU activation function, dropout with a probability of 0.25, and one output fully connected layer, which corresponds to the four underwater object classes. The Adam optimizer with a learning rate of 0.002 and a weight decay of 0.0001 was utilized to optimize the neural network. The network was trained for a maximum of 180 epochs with a batch size of 512. Early stopping was applied, which stopped the training when the validation accuracy did not improve for 20 consecutive epochs and restored the best model state according to the validation. The selected features were standardized using the statistics of the training set and were used as an input for the MLP.

4. Results and Discussion

This section presents an exhaustive analysis of the proposed framework for underwater image classification based on the assessment of the efficiency of each main component in the processing chain. First, the influence of the DSACM and TWSSL in enhancing spectral uniformity, structural integrity, and discriminatory features in underwater images was analyzed. The role of the hybrid DenseNet architecture in feature extraction was then examined in terms of its ability to encode complementary semantic information at multiple scales.

Next, the effect of the two-stage filtering technique and the Binary Weighted Grasshopper Optimization Algorithm (B-WGOA) in eliminating redundant features while retaining relevant data for classification is examined. A comparative analysis with respect to baseline architectures and prior works is provided in terms of performance measures such as accuracy,

precision, recall, and F1 score. Finally, the efficiency of the Label-Smoothed Weighted MLP classifier was evaluated based on stability and class-specific analysis.

4.1. Dataset Description and Representative Sample Visualization

Figure 5(a) shows examples of sample images from the UWI-2020 dataset [27], an underwater optical image benchmark created using ROV/AUV platforms for benthic habitat mapping under realistic underwater photography acquisition environments. This benchmark includes a wide range of environmental variations, such as illumination decay depending on water depth, particle-induced turbidity, wavelength-dependent color distortion, and backscatter noise, which jointly simulate real underwater vision operations. As depicted in Figure 5(a), the benchmark consists of four important categories of benthic biological organisms: Echinus, representing spiny echinoderms with radial symmetry; Holothurian, referring to long, soft-bodied sea cucumbers; Scallops, indicating fan-like bivalve molluscs anchored in sediments on the seabed; and Starfish, denoting multi-armed echinoderms with various morphologies. The diverse visual characteristics and environmental variations found in these samples make the benchmark an appropriate platform for validating enhancement-enabled deep classification models, including the proposed spectral-structural preprocessing method.

Figure 5(b) shows examples of sample images from the Marine_PULSE dataset [28], a side-scan sonar benchmark intended for detecting and analyzing subsea infrastructure and seabed morphology. In contrast to optical imaging, Marine_PULSE images include backscattered acoustic signals that encode both the geometric structure and material properties of underwater surfaces. As illustrated in Figure 5(b), the benchmark involves four types of structural sonar images: Engineering Platform (EP) for artificial offshore structures, Pipeline/Cable (PC) for linear subsea infrastructures, Sea Bed Surface (SBS) for natural seabed textures, and Underwater Residual Mound (URM) for irregular sediment accumulation. These sonar structural patterns offer complementary diversity to optical benthic imagery (UWI-2020), thus providing a second, modality-distinct dataset for evaluating the proposed classification framework.

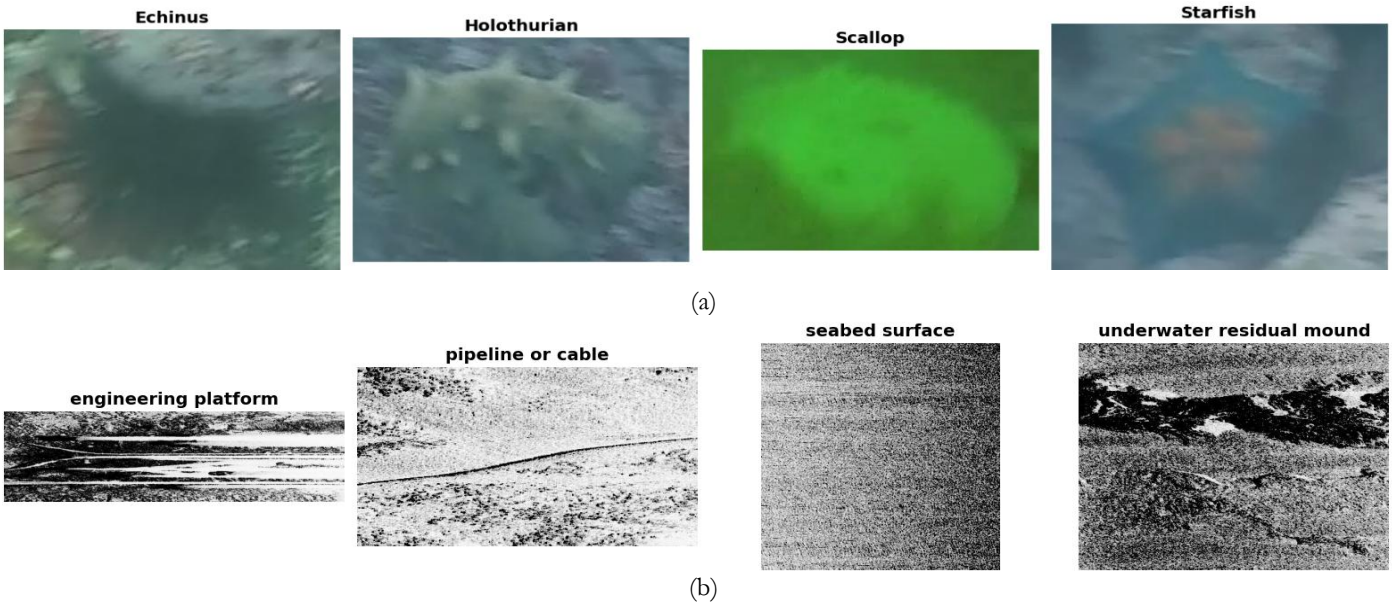
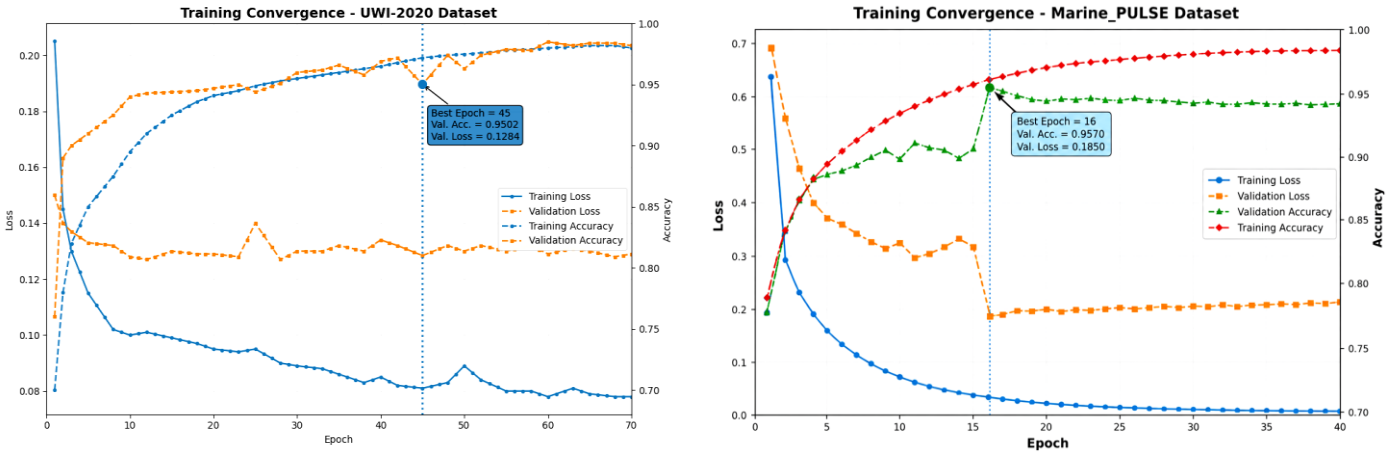


Figure 5. Representative Sample Images from the UWI-2020 Optical Benthic Dataset (**a**) and Marine_PULSE Side-Scan Sonar Dataset (**b**) Illustrating Class Diversity Across Marine Biological and Structural Categories

4.2. Training Convergence Analysis and Learning Behavior Evaluation

Figure 6(a) shows the training convergence curves of the developed classifier on the UWI-2020 dataset. The descending trajectory of the curve for training loss demonstrates that the model undergoes an accelerated learning process during early epochs, followed by a smooth decline towards the end of the training period, indicating optimization and learning of discriminative features. The corresponding training accuracy increased from initial epochs and stabilized at a certain level for the later epochs. Similar observations were made for the

accuracy on the validation set, which increased steadily until Epoch 45, when the best value of 95.02% was achieved, while the loss remained relatively stable with moderate fluctuations. The best validation accuracy was 95.02% at epoch 45. This validation result describes checkpoint selection during training and should be distinguished from the independent test-set metrics reported in Tables 5 and 6. Overall, the curves indicate convergence under the reported training conditions.



(a) (b)

Figure 6. Training Convergence Curves Illustrating the Convergence Behavior of the Proposed Framework on (**a**) the UWI-2020 Dataset and (**b**) the Marine_PULSE Side-Scan Sonar Dataset

Figure 6(b) shows the training behavior of the proposed classifier on the Marine_PULSE dataset. The training loss decreases substantially during the early epochs and then approaches zero, while training accuracy increases over the course of training. Validation accuracy reaches its highest value of 95.70% at epoch 16, after which the validation loss becomes relatively stable. This value is the best validation accuracy used for checkpoint selection; it is distinct from the independent test-set metrics reported in Table 12. The curves therefore describe the model's training and validation behavior on this dataset.

4.3. Performance Evaluation of the Proposed Framework on the UWI-2020 Underwater Optical Dataset

Table 5 presents the class-wise classification performance of the proposed DS-TSNet framework on the UWI-2020 underwater optical dataset. The results show consistently high precision, recall, and F1-score across all four benthic animal classes, with all values exceeding 0.94. Among the four classes, Starfish achieved the highest recall (0.9585) and F1-score (0.9541), indicating that the model correctly identified a relatively large proportion of Starfish samples. Scallops achieved the highest precision (0.9533), suggesting that predictions assigned to this class were comparatively less affected by false positives. Meanwhile, Echinus and Holothurian also demonstrated balanced precision and recall, with F1-scores of 0.9452 and 0.9425, respectively. These results indicate that the proposed framework maintains consistent classification performance across different benthic animal categories, despite variations in image appearance and morphological characteristics.

Table 5. Class-Wise Performance Metrics Derived from the Confusion Matrix for the UWI-2020 Underwater Dataset

Class	Precision	Recall	F1-Score	Support
Echinus	0.9409	0.9495	0.9452	218
Holothurian	0.9447	0.9404	0.9425	218
Scallops	0.9533	0.9401	0.9466	217
Starfish	0.9498	0.9585	0.9541	217
Macro Average	0.9472	0.9471	0.9471	—

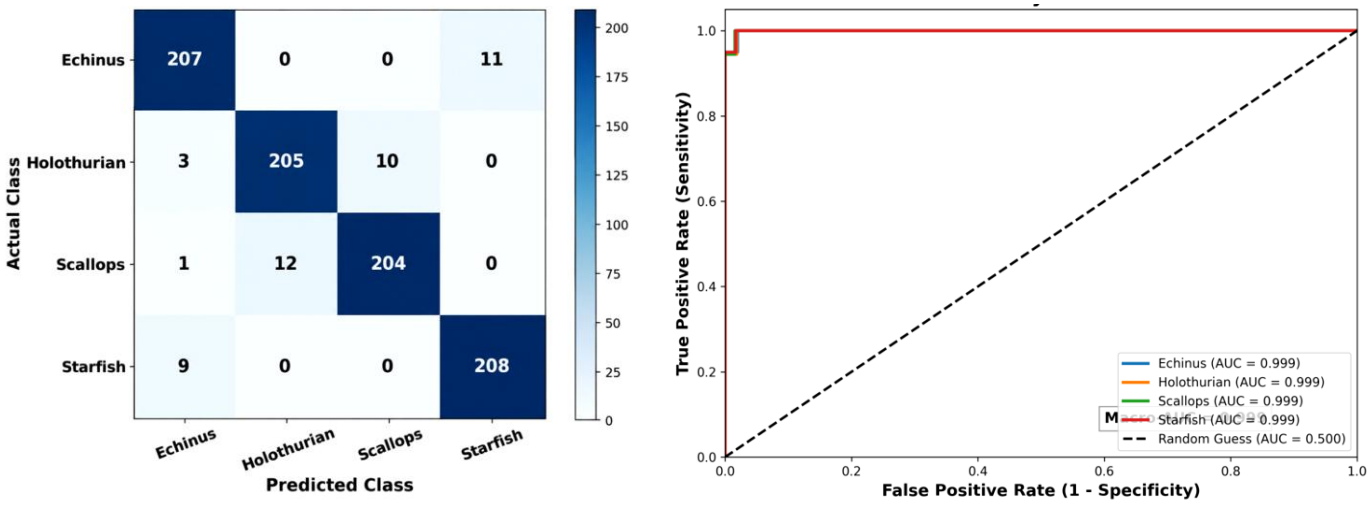
The overall performance metrics in Table 6 further demonstrate the classification capability of DS-TSNet. The framework achieved an accuracy of 94.71%, with macro-averaged precision, recall, and F1-score of 0.9472, 0.9471, and 0.9471, respectively. The close agreement among these metrics suggests that the model performs consistently across the four classes rather than relying predominantly on strong performance in a particular category. The corresponding error rate of 5.29% indicates that only a relatively small proportion of the evaluation samples were incorrectly classified.

Table 6. Global Performance Metrics of the Proposed Model on the UWI-2020 Dataset

Metric	Value
Overall Accuracy	94.71%
Macro Precision	0.9472
Macro Recall	0.9471
Macro F1-Score	0.9471
Average AUC (OvR)	0.9990
Error Rate	5.29%

The class-level prediction patterns and discriminative performance are further illustrated by the confusion matrix and one-vs-rest (OvR) ROC curves in Figures 7(a) and 7(b), respectively. The confusion matrix shows a strong concentration of predictions along the main diagonal, indicating that most samples were correctly assigned to their respective classes. The remaining misclassifications occurred primarily among the Echinus, Holothurian, and Starfish categories, which may be associated with similarities in their visual structures and textures. This observation is consistent with the slightly lower class-wise performance of Echinus and Holothurian compared with Starfish.

The ROC curves in Figure 7(b) provide further evidence of the model's discriminative capability, with an average OvR AUC of 0.9990. This high value indicates that the model can effectively distinguish each class from the remaining classes across different classification thresholds. Overall, the results from the class-wise metrics, global evaluation measures, confusion matrix, and ROC analysis provide complementary evidence of the classification performance of DS-TSNet on the UWI-2020 dataset.



(a) (b)

Figure 7. Performance Evaluation on the UWI-2020 Dataset: (**a**) Confusion Matrix and (**b**) One-vs-Rest ROC Curves Across All Classes

4.4. Spectral Compensation and Structural Stabilization Analysis Using DSACM–TWSSL Enhancement

The visual results in Figure 8, obtained using the DSACM and TWSSL modules, demonstrate successive changes in underwater image quality. In the first stage, DSACM addresses wavelength-dependent attenuation, particularly the reduction in red-channel intensity. This

results in color adjustment, contrast normalization, and more consistent scene illumination. The stage is intended to reduce the green-blue dominance commonly observed in underwater imagery.

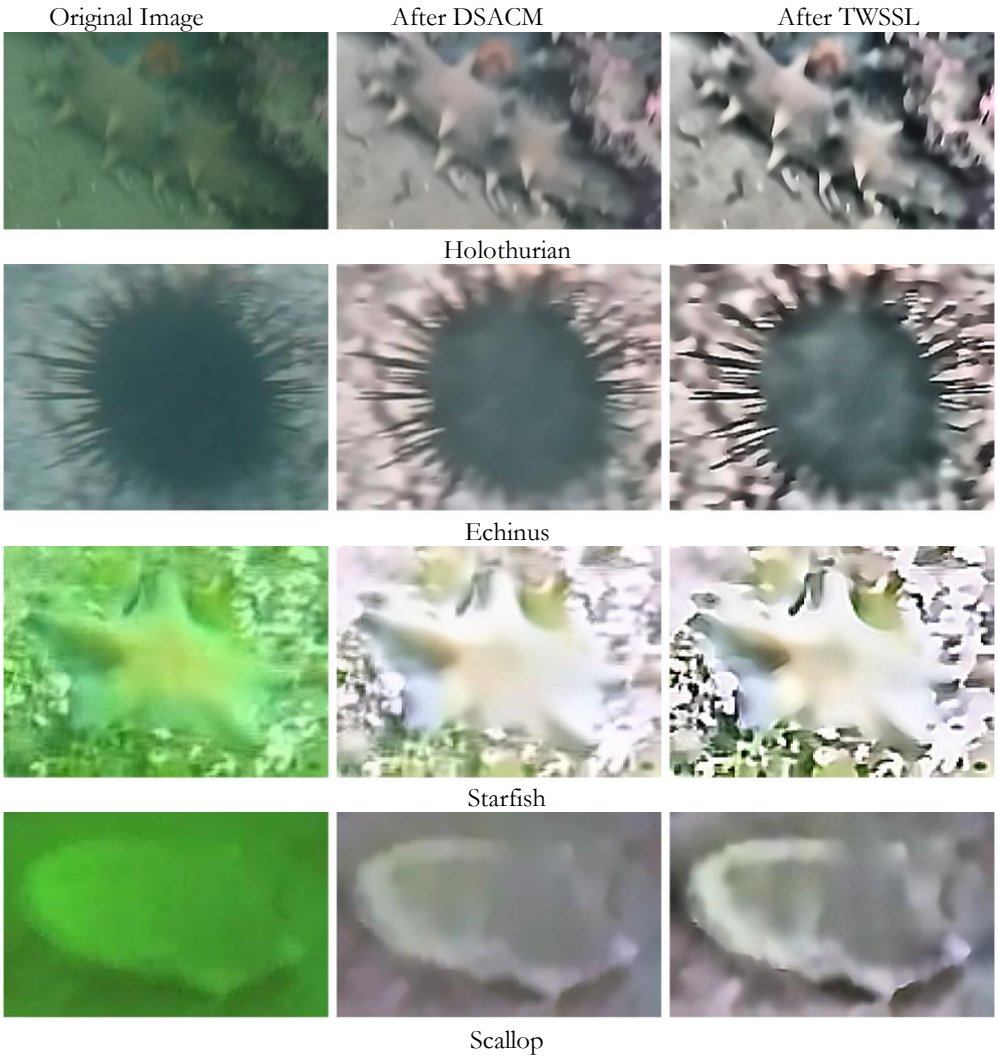


Figure 8. Visual Analysis of Depth-Aware Spectral Compensation and Turbidity-Weighted Structural Stabilization via DSACM–TWSSL Enhancement

In the second stage, the structural information is improved with turbidity reduction to minimize blur and backscatter while emphasizing important edges and textures that carry biological relevance. As a result of gradient-dependent structural weighting and filtering, the contours of the objects become sharper, and the contrast remains stable. Consequently, the images generated after applying the DSACM and TWSSL modules demonstrated increased semantic visibility compared to both the raw images and the images generated at the DSACM stage. The improved edge sharpness and contrast stabilization contribute to a more reliable extraction of feature representations during the later processing stages of the pipeline.

4.5. Comparative Evaluation with CNN and Optimization-Based Baseline Models

Table 7 presents a comparative evaluation of the proposed DS-TSNet framework against representative CNN architectures, including EfficientNet-B3, ResNet-50, and ResNet-101, as well as DenseNet-based configurations incorporating different feature selection strategies (DN201+WGOA and DN201+PSO+RFE). The proposed DS-TSNet achieved the highest accuracy of 94.71%, followed by DN201+WGOA (81.49%) and DN201+PSO+RFE (81.15%). The standalone CNN architectures achieved accuracies ranging from 76.10% to 78.50%. A similar performance trend was observed across precision, recall, F1-score, balanced accuracy, Cohen's kappa, and Matthews correlation coefficient (MCC). These results

indicate that DS-TSNet achieved better classification performance than the evaluated baseline configurations under the reported experimental settings.

Table 7. Comparative Performance Evaluation of the Proposed Model Against CNN and Optimization-Based Baselines on the UWI-2020 Dataset

Metric	Proposed Model	EfficientNet-B3	ResNet-50	ResNet-101	DN201+WGOA	DN201+PSO+RFE
Accuracy	0.9471	0.7850	0.7610	0.7760	0.8149	0.8115
Precision	0.9491	0.7882	0.7645	0.7793	0.8348	0.8320
Recall	0.9472	0.7846	0.7608	0.7759	0.8148	0.8114
F1-score	0.9474	0.7860	0.7621	0.7772	0.8182	0.8153
Weighted Precision	0.9492	0.7885	0.7649	0.7797	0.8350	0.8323
Weighted Recall	0.9471	0.7850	0.7610	0.7760	0.8149	0.8115
Weighted F1-score	0.9474	0.7863	0.7625	0.7775	0.8184	0.8155
Balanced Accuracy	0.9472	0.7846	0.7608	0.7759	0.8148	0.8114
Cohen's Kappa (κ)	0.9295	0.7132	0.6815	0.7024	0.7533	0.7487
Matthews CC	0.9300	0.7184	0.6869	0.7078	0.7584	0.7536

In particular, DS-TSNet achieved accuracy improvements of 13.22 and 13.56 percentage points over DN201+WGOA and DN201+PSO+RFE, respectively. Its Cohen's kappa and MCC values of 0.9295 and 0.9300, respectively, also indicate strong agreement between the predicted and actual class labels. Furthermore, the relatively high precision, recall, and F1-score values suggest that the proposed framework provides balanced classification performance across the evaluated classes. Overall, the results support the effectiveness of the integrated DS-TSNet configuration compared with the selected baseline models. However, these findings should be interpreted within the scope of the dataset and experimental settings used in this study. In particular, the reported performance differences demonstrate the comparative results on the evaluated dataset but do not, by themselves, establish statistical significance or consistency across different data partitions. Further evaluation through repeated experiments and statistical significance testing would be necessary to establish the reliability of these performance differences.

4.6. Ablation Study of Proposed DS-TSNet

The ablation study was conducted to examine the contributions of the main components of the proposed DS-TSNet framework and to investigate how their integration affects classification performance. As reported in Table 8, the hybrid DenseNet201–DenseNet121 architecture provides a better starting point than the single DenseNet121 backbone, suggesting that combining complementary feature representations can improve the characterization of underwater benthic organisms. The introduction of DSACM and TWSSL further improves the classification results. Notably, the configuration incorporating both modules performs better than those using either module individually. This observation suggests that spectral compensation and structural stabilization address complementary aspects of underwater image degradation and can jointly provide more informative representations for subsequent classification.

The results also highlight the contribution of feature selection to the overall framework. Incorporating variance and mutual-information filtering improves the performance of the combined preprocessing and hybrid feature-extraction configuration. The subsequent application of B-WGOA provides an additional improvement, indicating that selecting a more discriminative subset of features can benefit the classification process. The complete DS-TSNet framework with LSW-MLP achieves the best overall performance among the evaluated configurations. These findings suggest that the performance improvement is associated not only with the feature-extraction architecture but also with the integration of image enhancement and feature-selection strategies. Nevertheless, the contribution of each component should be interpreted within the evaluated experimental configurations rather than as evidence of independent or statistically significant effects.

To further investigate the role of feature filtering, Table 9 compares variance filtering, mutual-information filtering, and their combination while maintaining the same underlying feature representation. Both individual filtering strategies improve the classification results

compared with the configuration without filtering. MI filtering produces slightly better overall results than variance filtering when considered individually, suggesting that relevance-based feature selection is useful for retaining features associated with class discrimination. The combined strategy achieves the highest accuracy and precision among the evaluated filtering configurations. However, its recall and macro-F1 score are lower than those obtained with the individual filtering approaches. This indicates that combining the two strategies does not uniformly improve all performance measures. Rather, the results reveal a trade-off between improving overall prediction accuracy and maintaining balanced classification performance across classes.

Table 8. Ablation Study Showing Progressive Performance Improvements of DS-TSNet Components on the UWI-2020 Dataset

Configuration	Accuracy	Precision	Recall	Macro F1-score	MCC
DenseNet121 Baseline(BL)	0.781	0.788	0.776	0.774	0.689
Hybrid DenseNet201–121(HYB)	0.802	0.806	0.798	0.799	0.731
HYB+ DSACM	0.842	0.847	0.836	0.836	0.742
HYB+ TWSSL	0.861	0.866	0.854	0.854	0.768
HYB+ DSACM + TWSSL	0.903	0.907	0.897	0.897	0.824
HYB+DSACM + TWSSL+Variance + MI Filtering	0.926	0.929	0.921	0.921	0.871
HYB+DSACM + TWSSL+Variance + MI Filtering+ B-WGOA Feature Selection	0.939	0.941	0.936	0.936	0.907
Full DS-TSNet with LSW-MLP Classifier (Proposed)	0.947	0.949	0.947	0.947	0.930

Table 9. Ablation Study of Feature Filtering Strategies on the UWI-2020 Dataset

Configuration	Accuracy	Precision	Recall	Macro F1-score
DSACM + TWSSL + Hybrid DenseNet201–121 (Without Filtering — Baseline)	0.903	0.907	0.897	0.897
DSACM + TWSSL + Hybrid DenseNet201–121 + Variance Filtering Only	0.918	0.915	0.924	0.926
DSACM + TWSSL + Hybrid DenseNet201–121 + MI Filtering Only	0.92	0.918	0.928	0.926
DSACM + TWSSL + Hybrid DenseNet201–121 + Variance + MI Filtering	0.926	0.929	0.921	0.921

The ablation results support the effectiveness of integrating image enhancement, hybrid feature extraction, and feature-selection strategies within DS-TSNet. The findings also indicate that the contribution of feature filtering depends on the evaluation metric, highlighting the importance of considering multiple performance measures when selecting an appropriate feature-reduction strategy. Further analysis of the computational cost and feature dimensionality would provide additional insight into the practical benefits of the proposed pipeline.

4.7. Computational Complexity and Feature Selection Analysis

Table 10 shows the computational performance and characteristics of the proposed DS-TSNet framework on the UWI-2020 dataset. The experiments were conducted on a Tesla T4 GPU with CUDA 12.4, with 3,480 training and 870 independent evaluation images distributed across four marine organism classes. The 3,480 training samples were split into 3,132 and 348 training and internal validation samples, respectively, with the 870 evaluation samples reserved for the final testing stage.

The timing results for the DSACM and TWSSL feature extractors showed that the former took 93.90 ms on average per image, while the latter took only 10.10 ms, for a combined processing time of 104.00 ms per image. The hybrid feature vector of size 2,944 was obtained in 190.641 s for the training set and 78.217 s for the evaluation set. The subsequent feature selection pipeline reduced the dimensionality of the data from 2,944 to 2,400 based on the variance criterion and further down to 1,200 features based on the mutual information between the variables.

The final B-WGOA stage acted on the 1,200-dimensional data and stopped after five of the maximally allowed 10 iterations, with the fitness value remaining at 0.111793. It selected 579 features out of the available 1,200, taking 2,343.923 s in total for the computation. Thus, the entire pipeline reduced the original 2,944 features by 2,365 dimensions or 80.33% of the

initial value, leaving only 19.67% of the features for the downstream classification stage. In this manner, the proposed approach both drastically reduces the dimensionality of the original hybrid feature vector while also allowing for computational profiling and early stopping of the optimization process in the case of insufficient fitness improvement.

Table 10. Computational Profiling and Progressive Feature Reduction of the Proposed DS-TSNet Framework on the UWI-2020 Dataset

Category	Parameter / Metric	Value
Hardware	Device	CUDA
	GPU	Tesla T4
	CUDA version	12.4
Dataset	Training images	3,480
	Evaluation images	870
	Total images	4,350
	Number of classes	4
	Classes	Echinus, Holothurian, Scallop, Starfish
Data split	Classifier training samples	3,132
	Internal validation samples	348
	Official evaluation samples	870
DSACM + TWSSL timing	Images profiled	100
	DSACM mean time/image	93.90 ms
	DSACM standard deviation	229.10 ms
	TWSSL mean time/image	10.10 ms
	TWSSL standard deviation	25.42 ms
	Combined mean time/image	104.00 ms
	Combined standard deviation	254.41 ms
Hybrid features	Training feature dimensions	2,944
	Evaluation feature dimensions	2,944
	Training extraction time	190.641 s
	Evaluation extraction time	78.217 s
Variance filtering	Input dimensions	2,944
	Variance-selected dimensions	2,400
	Variance filtering time	0.1976 s
Mutual information	Input dimensions	2,400
	Final MI dimensions	1,200
	MI filtering time	30.9993 s
B-WGOA	Input dimensions	1,200
	Maximum iterations	10
	Iterations executed	5
	Initial fitness	0.111793
	Final/best fitness	0.111793
	Features during optimization	589 → 603 → 575 → 573 → 567
	Final selected dimensions	579
	B-WGOA execution time	2,343.923 s
Overall feature reduction	Original dimensions	2,944
	Final dimensions	579
	Features removed	2,365
	Overall dimensionality reduction	80.33%
	Features retained	19.67%

The feature selection is performed offline, only the retained feature indices are applied during inference. Thus, the framework achieves substantial feature compactness while concentrating its optimization cost in the offline stage. Classifier training and complete end-to-end inference times are distinguished from the reported component-wise measurements.

4.8. Comparative Analysis with Existing Methods

Table 11 presents a comparative analysis of the proposed DS-TSNet framework against several underwater image classification approaches reported in the literature, including feature-based methods, segmentation-based techniques, and transfer learning models. Among the evaluated approaches, the Flow Direction Algorithm combined with an Extreme Learning Machine (FDA-ELM) [29] achieved an accuracy of 94.59%, demonstrating competitive classification performance. However, its reliance on manually designed directional features may limit its ability to capture more complex visual representations under varying underwater imaging conditions. Similarly, the Dual-Attention U-Net (DAU-Net) [30], achieved a precision of 95.26% and a recall of 95.17%, with an overall accuracy of 93.95%. Although these results indicate strong performance in the reported study, the segmentation-oriented design of DAU-Net differs from the integrated spectral compensation, structural enhancement, and feature-selection strategy adopted in the proposed framework.

Conventional feature-based and transfer learning approaches reported in [20] achieved accuracies of 84.31% for Bag-of-Features with SIFT and SVM (BoF+SIFT) and 92.87% for VGG16, respectively. In comparison, the proposed DS-TSNet achieved the highest accuracy among the methods listed in Table 11, reaching 94.71%, with precision, recall, and F1-score values of 94.91%, 94.72%, and 94.74%, respectively. The relatively balanced values across these metrics suggest consistent classification performance across the evaluated classes. Notably, the accuracy improvement over FDA-ELM was 0.12 percentage points, while the difference was 0.76 percentage points compared with DAU-Net. These results indicate that DS-TSNet provides competitive performance relative to the selected approaches, although the observed differences should be interpreted in the context of the experimental protocols and datasets used in the respective studies.

The performance of DS-TSNet can be considered in relation to its integrated design, which combines spectral compensation through DSACM, turbidity-related structural enhancement through TWSSL, hybrid feature extraction using DenseNet architectures, and optimization-guided feature selection using the B-WGOA algorithm. These components are intended to improve the representation and selection of discriminative features for underwater image classification. Nevertheless, since the comparative results are drawn from previously reported studies, they do not establish a definitive performance advantage under identical experimental conditions.

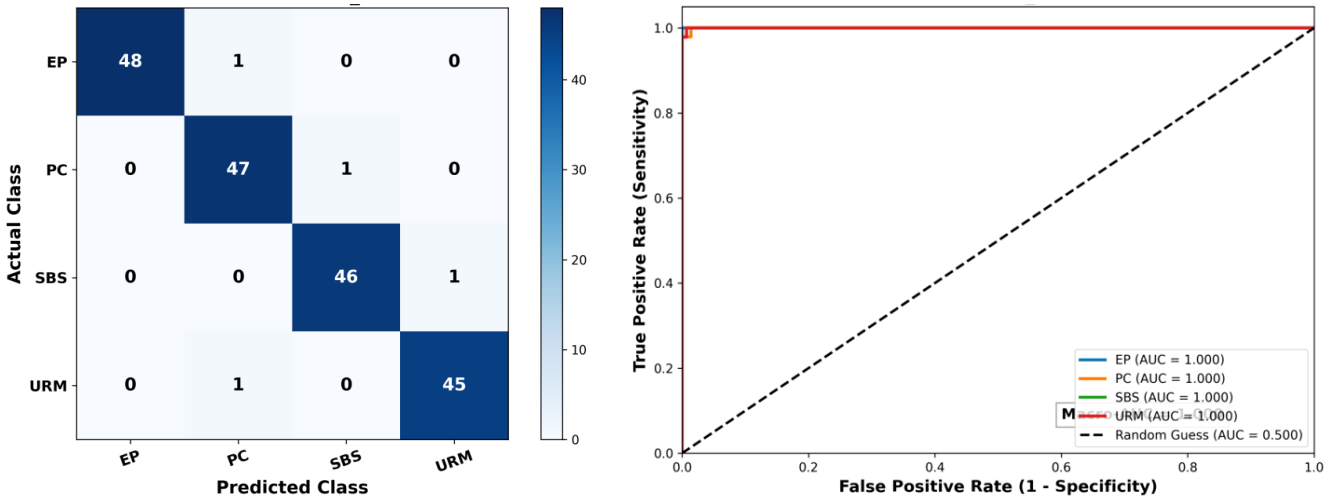
Table 11. Comparative Performance of the Proposed DS-TSNet Framework and Previously Reported Underwater Image Classification Methods.

Ref	Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
[29]	Flow Direction Algorithm + Extreme Learning Machine (FDA-ELM)	94.59	94.16	93.41	93.77
[30]	Dual-Attention U-Net (DAU-Net)	93.95	95.26	95.17	95.24
[20]	Bag-of-Features + SIFT (BoF+SIFT (SVM))	84.31	—	—	—
[20]	Transfer Learning (VGG16)	92.87	—	—	—
Ours	DS-TSNet	94.71	94.91	94.72	94.74

4.9. Evaluation on the Marine_PULSE Dataset

To evaluate the proposed framework on a second dataset, additional experiments were conducted using the Marine_PULSE side-scan sonar dataset. The framework was trained and tested on Marine_PULSE using its corresponding class labels, while retaining the methodological pipeline used for the UWI-2020 optical dataset. Thus, this experiment evaluates the framework separately on an acoustic dataset; it does not constitute a direct cross-dataset transfer test from UWI-2020 to Marine_PULSE. Table 12 reports an accuracy, macro precision, macro recall, and macro F1-score for Marine_PULSE. These results describe performance

on the Marine_PULSE evaluation split and should not be interpreted as evidence of transfer between optical and acoustic modalities.



(a) (b)

Figure 9. Evaluation Performance of the Proposed DS-TSNet Framework on the Marine_PULSE Dataset: (**a**) Confusion Matrix and (**b**) One-vs-Rest ROC Curves Across EP, PC, SBS, and URM Classes

Furthermore, the confusion matrix in Figure 9(a) shows that most predictions lie on the diagonal, indicating a small number of classification errors. The ROC curves in Figure 9(b) report an AUC of 1.000 for the evaluated classes. These results indicate strong class separability within the Marine_PULSE evaluation setting. They do not, by themselves, establish transferability between optical underwater images and sonar data.

Table 12. Classification Performance of the Proposed DS-TSNet Framework on the Marine_PULSE Dataset

Metric	Value
Overall Accuracy	97.89%
Macro Precision	0.9786
Macro Recall	0.9781
Macro F1-Score	0.9783
Average AUC (OvR)	1.0000
Error Rate	2.11%

5. Limitations of the Study

Although the results obtained on UWI-2020 (optical imagery) and Marine_PULSE (acoustic imagery) are promising, several limitations should be acknowledged before considering practical underwater applications. First, the proposed approach was evaluated using publicly available benchmark datasets, which may not fully represent real-world underwater conditions involving variations in illumination, water composition, sensor characteristics, viewing angles, and environmental factors. Furthermore, the relatively controlled class distributions and data acquisition settings of these datasets may not adequately capture the diversity and complexity of operational underwater environments, where environmental conditions and acquisition characteristics can vary considerably.

Second, although the study investigates the potential of DS-TSNet for joint optical-acoustic analysis, the experiments were conducted using separate optical and acoustic datasets rather than a single synchronized multimodal dataset. Consequently, the effectiveness of jointly processing simultaneously acquired optical and acoustic information in practical underwater scenarios remains to be validated. In particular, the potential benefits of cross-modal information sharing and complementary feature extraction under real-world acquisition conditions have not yet been fully established.

Third, the environmental condition estimation in the current framework relies primarily on spectral and structural information extracted from the input images. Specifically, the DSACM and TWSSL components employ channel-wise compensation mechanisms and image-gradient-based proxies for turbidity estimation rather than explicit depth information, direct turbidity measurements, or physical attenuation properties. Consequently, these image-derived indicators may not fully characterize complex underwater conditions, particularly in environments with severe turbidity, varying water depths, or substantial light attenuation. Incorporating depth estimates from dedicated sensors, direct turbidity measurements, or physical attenuation models could provide more reliable environmental information and potentially improve the accuracy and interpretability of the proposed approach.

Finally, the optimization procedure involves stochastic search, meaning that the selected feature subset may be influenced by the initial population and random initialization. Although multiple restarts were employed during fitness evaluation, further experiments are needed to quantify the variability of the selected feature subsets and determine how such variations affect the performance of the subsequent MLP classifier. Moreover, the overall pipeline, which combines image preprocessing, dual DenseNet feature extraction, filter-based feature selection, B-WGOA optimization, and MLP classification, introduces considerable computational and memory overhead. This complexity may limit its applicability in resource-constrained underwater systems. Therefore, further investigation into computational optimization, feature dimensionality reduction, and model simplification would be beneficial to improve deployment efficiency, particularly for lightweight embedded platforms.

6. Conclusions

This study proposed DS-TSNet, an integrated framework for underwater image classification that addresses the challenges associated with spectral degradation, turbidity-related structural distortion, and high-dimensional feature representations. The experimental findings demonstrate the effectiveness of combining spectral compensation and structural stabilization with hybrid deep feature extraction and optimization-guided feature selection. The component-wise evaluations further indicate that these stages contribute to improving the discriminative quality of the feature representation and the overall classification performance. On the UWI-2020 optical dataset, DS-TSNet achieved an accuracy of 94.71%, while a separate evaluation on the Marine_PULSE acoustic dataset yielded an accuracy of 97.89%. These results demonstrate the framework's ability to achieve promising classification performance across two distinct underwater imaging datasets.

Nevertheless, the results should be interpreted within the scope of the evaluated datasets and experimental protocols. Since the framework was trained and evaluated independently on each dataset, its performance does not establish cross-modal knowledge transfer or generalization between optical and acoustic imaging modalities. Moreover, the current evaluation does not fully represent the variability of real-world underwater environments. Future research will therefore investigate DS-TSNet using synchronized multimodal marine datasets and more diverse field-acquired data to assess its robustness under varying environmental and acquisition conditions. Further work will also explore model compression and computational optimization to facilitate deployment on resource-constrained edge devices within autonomous underwater monitoring systems.

Author Contributions: Conceptualization: M. S. and J. B. S.; Methodology: M. S. and J. B. S.; Software: M. S.; Validation: M. S.; Formal analysis: M. S.; Investigation: M. S. and J. B. S.; Resources: M. S. and J. B. S.; Data curation: M. S.; Writing—original draft preparation: M. S.; Writing—review and editing: M. S. and J. B. S.; Visualization: M. S. and J. B. S.; Supervision: J. B. S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets used in this study can be downloaded/accessed here: Nusrath Ibrahim. (2025). Underwater_Image_Dataset [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/12732449 and Nusrath Ibrahim. (2024). Marine Pulse [Data set]. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/10342860.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] E. Prasetyo, N. Suciati, and C. Fatichah, "Multi-level residual network VGGNet for fish species classification," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 8, pp. 5286–5295, Sep. 2022, doi: 10.1016/j.jksuci.2021.05.015.

[2] Mehrunnisa, M. Leszczuk, D. Juszka, and Y. Zhang, "Improved Binary Classification of Underwater Images Using a Modified ResNet-18 Model," *Electronics*, vol. 14, no. 15, p. 2954, Jul. 2025, doi: 10.3390/electronics14152954.

[3] S. Liawatimena, E. Abdurachman, A. Trisetyarso, A. Wibowo, M. K. Ario, and I. S. Edbert, "Fish Classification System Using YOLOv3-ResNet18 Model for Mobile Phones," *CommIT (Communication Inf. Technol. J.*, vol. 17, no. 1, pp. 71–79, Mar. 2023, doi: 10.21512/commit.v17i1.8107.

[4] P. Pachaiyappan, G. Chidambaram, A. Jahid, and M. H. Alsharif, "Enhancing Underwater Object Detection and Classification Using Advanced Imaging Techniques: A Novel Approach with Diffusion Models," *Sustainability*, vol. 16, no. 17, p. 7488, Aug. 2024, doi: 10.3390/su16177488.

[5] H. Xin and L. Li, "Arbitrary Style Transfer With Fused Convolutional Block Attention Modules," *IEEE Access*, vol. 11, pp. 44977–44988, 2023, doi: 10.1109/ACCESS.2023.3273949.

[6] H. Zhang, R. He, and W. Fang, "An Underwater Image Enhancement Method Based on Diffusion Model Using Dual-Layer Attention Mechanism," *Water*, vol. 16, no. 13, p. 1813, Jun. 2024, doi: 10.3390/w16131813.

[7] R. Liu, Z. Jiang, S. Yang, and X. Fan, "Twin Adversarial Contrastive Learning for Underwater Image Enhancement and Beyond," *IEEE Trans. Image Process.*, vol. 31, pp. 4922–4936, 2022, doi: 10.1109/TIP.2022.3190209.

[8] L. Peng, C. Zhu, and L. Bian, "U-Shape Transformer for Underwater Image Enhancement," *IEEE Trans. Image Process.*, vol. 32, pp. 3066–3079, 2023, doi: 10.1109/TIP.2023.3276332.

[9] X. Wang, G. Xue, S. Huang, and Y. Liu, "Underwater Object Detection Algorithm Based on Adding Channel and Spatial Fusion Attention Mechanism," *J. Mar. Sci. Eng.*, vol. 11, no. 6, p. 1116, May 2023, doi: 10.3390/jmse11061116.

[10] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated Residual Transformations for Deep Neural Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 5987–5995. doi: 10.1109/CVPR.2017.634.

[11] V. Lopez-Vazquez, J. M. Lopez-Guede, D. Chatzievangelou, and J. Aguzzi, "Deep learning based deep-sea automatic image enhancement and animal species classification," *J. Big Data*, vol. 10, no. 1, p. 37, Mar. 2023, doi: 10.1186/s40537-023-00711-w.

[12] S. D *et al.*, "Enhanced deep learning models for automatic fish species identification in underwater imagery," *Heliyon*, vol. 10, no. 15, p. e35217, Aug. 2024, doi: 10.1016/j.heliyon.2024.e35217.

[13] S. Kay, K. Adhikari, and B. Tang, "A Nonparametric Data-Driven Classifier Based on the Cumulant Generating Function," *IEEE Trans. Signal Process.*, vol. 73, pp. 519–533, 2025, doi: 10.1109/TSP.2025.3525951.

[14] Y. Hwang, G. Kim, S. Shin, W. Hong, and Y. Choo, "Attention-Based Complementary Learning for Active Target Classification With Limited Sonar Data," *IEEE Access*, vol. 12, pp. 79787–79801, 2024, doi: 10.1109/ACCESS.2024.3409829.

[15] B. Muñoz and G. Troni, "Learning the Ego-Motion of an Underwater Imaging Sonar: A Comparative Experimental Evaluation of Novel CNN and RCNN Approaches," *IEEE Robot. Autom. Lett.*, vol. 9, no. 3, pp. 2072–2079, Mar. 2024, doi: 10.1109/LRA.2024.3352357.

[16] Z. Feng, P. Zhang, and X. Huo, "Underwater Target Classification Based on Feature Fusion and Gene Encoding of CNN-BIGRU-Attention," *IEEE Access*, vol. 11, pp. 139546–139556, 2023, doi: 10.1109/ACCESS.2023.3341499.

[17] J. Liu, S. Liu, S. Xu, and C. Zhou, "Two-Stage Underwater Object Detection Network Using Swin Transformer," *IEEE Access*, vol. 10, pp. 117235–117247, 2022, doi: 10.1109/ACCESS.2022.3219592.

[18] S. C. Pravin, G. Rohith, V. Kiruthika, E. Manikandan, S. Methelesh, and A. Manoj, "Underwater Animal Identification and Classification Using a Hybrid Classical-Quantum Algorithm," *IEEE Access*, vol. 11, pp. 141902–141914, 2023, doi: 10.1109/ACCESS.2023.3343120.

[19] J. de C. V. Fernandes, N. N. de Moura Junior, and J. M. de Seixas, "Deep Learning Models for Passive Sonar Signal Classification of Military Data," *Remote Sens.*, vol. 14, no. 11, p. 2648, Jun. 2022, doi: 10.3390/rs14112648.

[20] G. Huo, Z. Wu, and J. Li, "Underwater Object Classification in Sidescan Sonar Images Using Deep Transfer Learning and Semisynthetic Training Data," *IEEE Access*, vol. 8, pp. 47407–47418, 2020, doi: 10.1109/ACCESS.2020.2978880.

[21] Y. Hwang, G. Kim, W. Hong, and Y. Choo, "Effective Data Augmentation for Active Sonar Classification Using Attention-Based Complementary Learning With Uncertainty Measure," *IEEE Access*, vol. 13, pp. 55670–55681, 2025, doi: 10.1109/ACCESS.2025.3555420.

[22] C. Barat and R. Phlypo, "A Fully Automated Method to Detect and Segment a Manufactured Object in an Underwater Color Image," *EURASIP J. Adv. Signal Process.*, vol. 2010, no. 1, p. 568092, Dec. 2010, doi: 10.1155/2010/568092.

[23] M. A. Abdel-Moneim, E.-S. M. El-Rabaie, F. E. A. El-Samie, K. Ramadan, K. F. Ramadan, and N. Abdel-Salam, "CNN-based modulation classification using image processing techniques for OFDM underwater acoustic communication system," *J. Electr. Syst. Inf. Technol.*, vol. 12, no. 1, p. 70, Sep. 2025, doi: 10.1186/s43067-025-00263-4.

[24] Y. Steiniger, D. Kraus, and T. Meisen, "Survey on deep learning based computer vision for sonar imagery," *Eng. Appl. Artif. Intell.*, vol. 114, p. 105157, Sep. 2022, doi: 10.1016/j.engappai.2022.105157.

[25] P. Prasenan and C. D. Suriyakala, "Fish species classification using a collaborative technique of firefly algorithm and neural network," *EURASIP J. Adv. Signal Process.*, vol. 2022, no. 1, p. 116, Dec. 2022, doi: 10.1186/s13634-022-00950-8.

[26] X. Qin, X. Luo, Z. Wu, and J. Shang, "Optimizing the Sediment Classification of Small Side-Scan Sonar Images Based on Deep Learning," *IEEE Access*, vol. 9, pp. 29416–29428, 2021, doi: 10.1109/ACCESS.2021.3052206.

[27] N. Ibrahim, "Underwater_Image_Dataset," *Kaggle*, 2025. https://www.kaggle.com/datasets/khajaahmed1/underwater-image-dataset
[28] N. Ibrahim, "Marine_Pulse," *Kaggle*, 2024. https://www.kaggle.com/datasets/khajaahmed1/marine-pulse
[29] J. Yang, M. Cai, X. Yang, and Z. Zhou, "Underwater Image Classification Algorithm Based on Convolutional Neural Network and Optimized Extreme Learning Machine," *J. Mar. Sci. Eng.*, vol. 10, no. 12, p. 1841, Dec. 2022, doi: 10.3390/jmse10121841.
[30] Y. Ren, X. Li, X. Yang, and H. Xu, "Development of a Dual-Attention U-Net Model for Sea Ice and Open Water Classification on SAR Images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2021.3058049.