E-ISSN: 2963-4946
Vol. 1 No. 3 March 2022, pp. 140-155
DOI: 10.59888/ajosh.v4i10.787

Asian Journal of Social and Humanities

Marine Image Classification Using EfficientNetV2, Focal Loss, and XAI

Wahyu Bontot
Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) IKMI Cirebon, Indonesia
Corresponding Author: bontotwahyu8@gmail.com

Keywords:
marine images; class imbalance; EfficientNetV2; focal loss; XAI

This is an open access article under the CC BY-SA license.

Copyright © 2026 by Author. Published by Pelopor Publikaasi Akademika

Abstract

Background: Marine ecosystems require fast and accurate monitoring systems, yet the analysis of underwater images faces complex environmental conditions and class imbalance.
Objective: This study proposes the integration of an EfficientNetV2-based deep learning classifier with Adaptive Underwater Image Enhancement (AUIE), focal loss, and Explainable Artificial Intelligence (XAI) to improve marine image classification.
Method: The dataset consists of 22 classes of marine organisms divided into training, validation, and test sets with a ratio of 80:10:10. AUIE combines Contrast Limited Adaptive Histogram Equalization on the lightness channel with colour balance correction, and focal loss with $\gamma = 2.0$ replaces standard cross-entropy. This study used a quantitative experimental approach with four experimental scenarios.
Results: An ablation study of four scenarios showed that the baseline reached 79.47% accuracy, whereas the combination of AUIE and focal loss produced a validation loss of 0.5503 and an accuracy of 76.32%, recovering part of the accuracy lost when AUIE was used with cross-entropy. Confusion matrix inspection indicated fewer misclassifications of camouflaged minority classes, and Grad-CAM showed attention concentrated on the morphological features of the organisms after enhancement.
Conclusion: The study contributes to robust and transparent marine biodiversity monitoring systems, while longer training and per-class evaluation are still required.

INTRODUCTION

Marine ecosystems are a crucial component in maintaining the balance of the global environment. Modern marine monitoring relies heavily on underwater images captured by cameras or uncrewed vehicles such as remotely operated vehicles (ROVs) and autonomous underwater vehicles (AUVs). However, the very large volume of visual data demands automation using computer vision and deep learning technologies, such as convolutional neural networks (CNNs). Goodwin et al., (2022) showed that deep learning has opened new possibilities across marine ecology, from species identification to abundance estimation, but also noted that its adoption depends on the availability of representative training data and on the trust of domain experts in model outputs.

Manual annotation of underwater imagery is slow and requires taxonomic expertise, and the number of images collected by camera stations and survey vehicles increasingly

exceeds what human experts can inspect. Automated classifiers can reduce this burden by pre-labelling images, prioritizing uncertain cases for expert review, and enabling near-real-time summaries of species occurrence (Saleh et al., 2022). For such systems to be useful in practice, however, they must remain reliable under the degraded visual conditions typical of underwater scenes, must not neglect rare organisms, and must provide outputs that experts can verify.

Although CNNs excel at feature extraction, the classification of underwater images faces major challenges arising from complex environmental conditions, such as uneven illumination, noise, low contrast, and colour casts caused by the wavelength-dependent absorption and scattering of light in water (Raveendran et al., 2021). Saleh et al. (2022) reviewed computer vision and deep learning for fish classification in underwater habitats and identified image quality, background complexity, and limited labelled data as recurring obstacles. These degradations reduce the visibility of the morphological details that distinguish one organism from another, especially when organisms are camouflaged against coral reefs or sand.

A second challenge is class imbalance, that is, the unequal number of samples across classes. Marine datasets are generally dominated by majority classes, which causes models to lose the ability to recognize minority species or objects that are precisely important for conservation. Johnson and Khoshgoftaar (2019) surveyed deep learning with class imbalance and grouped the countermeasures into data-level methods, such as resampling, and algorithm-level methods, such as cost-sensitive or re-weighted loss functions. In the specific context of taxon identification, Langenkämper et al. (2019) compared mitigation strategies and showed that their effectiveness depends on the data and the evaluation metric used.

Image enhancement is a common response to the first challenge. Classical methods include histogram-based contrast enhancement and colour correction, such as the colour balance and fusion approach of Ancuti et al., (2018), whereas recent methods use deep networks trained to restore underwater images (Raveendran et al., 2021). Liu et al. (2020) argued that enhancement should be considered together with the downstream perception task, because an image that looks better to a human observer does not necessarily lead to better recognition by a model. Lopez-Vazquez et al. (2020), for example, integrated CLAHE-based enhancement and bilateral filtering into a pipeline for classifying deep-sea fauna at a cabled observatory and reported an accuracy of 76.18% and an area under the curve (AUC) of 87.59% on an independent test set.

Among modern CNN architectures, EfficientNetV2 offers a favourable balance between accuracy and training efficiency (Tan & Le, 2021). Kyathanahally et al. (2021) compared several deep learning models for the classification of aquatic zooplankton images and found that EfficientNet models were among the best performers, with EfficientNetB7 achieving the highest F1-score (about 90.0%) among the single image models. These results make the EfficientNet family, including EfficientNetV2, a reasonable backbone for marine image classification, although performance on rare classes and under strong image degradation remains more difficult. Kyathanahally et al.

(2021) reported, for instance, that their best ensemble reached 97.9% accuracy but a macro-averaged F1-score of 92.7% on data in which the most common class had about 300 times more images than the rarest class, which shows how strongly class imbalance affects reported performance. Table 1 summarizes the empirical studies most closely related to this work.

Table 1. Related empirical studies on marine image classification

Study	Data	Method	Reported result
Lopez-Vazquez et al. (2020)	Time-lapse images from the LoVe cabled observatory at about 260 m depth, Norway (13 classes)	CLAHE-based enhancement and bilateral filtering followed by machine and deep learning classifiers	Accuracy 76.18%; AUC 87.59% (best deep neural network, independent test set)
Kyathanahally et al. (2021)	Lake zooplankton images from an in situ plankton camera (35 classes; 17,943 images)	Comparison of deep models; EfficientNet models and ensembles	Accuracy 97.9% and macro F1 92.7% (best ensemble); F1 about 90.0% (EfficientNetB7)
This study	Sea Animals Image Dataset; 22 classes; 13,297 images	EfficientNetV2-based classifier with AUIE, focal loss, and Grad-CAM	Accuracy 74.96%–79.47% across four ablation scenarios

Source: Data Processed

A third concern is transparency. Deep learning models are often described as black boxes, which makes it difficult for marine scientists to judge whether a prediction is based on the organism or on incidental background features. Explainable Artificial Intelligence (XAI) addresses this problem by providing explanations of model behaviour; Barredo Arrieta et al. (2020) organized the many available XAI methods into a taxonomy, and Linardatos et al. (2020) reviewed these interpretability methods and their implementations. Gradient-weighted Class Activation Mapping (Grad-CAM) is one of the most widely used methods for CNNs because it produces a class-specific heatmap from the gradients flowing into the last convolutional layer (Selvaraju et al., 2020).

Table 2 summarizes how the three challenges are addressed in the literature and in the present study. This study combines EfficientNetV2, Adaptive Underwater Image Enhancement (AUIE), focal loss, and Grad-CAM within a single 2 × 2 ablation framework. The novelty does not lie in proposing a new enhancement, loss, or explainability algorithm; rather, it lies in examining how the preprocessing and loss-function choices perform separately and jointly in a 22-class marine-image classification task, while considering model explanations as an interpretive component. The available studies reviewed in this manuscript do not establish whether the same combination has

been evaluated on this dataset; therefore, a claim of being the first study using this dataset is not made.

Table 2. Challenges in marine image classification and the approaches considered

Challenge	Typical approaches in the literature	Approach in this study
Image degradation (low contrast, colour cast, noise)	Histogram-based enhancement, colour correction and fusion, deep enhancement networks (Ancuti et al., 2018; Islam et al., 2020; Raveendran et al., 2021)	AUIE: CLAHE on the LAB lightness channel combined with colour balance correction
Class imbalance	Resampling, cost-sensitive and re-weighted losses (Johnson & Khoshgoftaar, 2019; Langenkämper et al., 2019)	Focal loss with $\gamma = 2.0$ instead of cross-entropy
Limited transparency	Post hoc XAI methods such as saliency and class activation maps (Barredo Arrieta et al., 2020; Linardatos et al., 2020)	Grad-CAM heatmaps on the trained classifier

Source: Data Processed

The objectives of this study are to: (1) evaluate the effect of AUIE preprocessing and focal loss, separately and in combination, on the classification of 22 classes of marine organisms through an ablation study; and (2) examine, through Grad-CAM, whether the proposed configuration directs the attention of the model to the morphological features of the organisms.

The expected contribution is an integrated and transparent classification pipeline that can support the development of marine biodiversity monitoring systems. Its practical benefit is to provide an empirical basis for selecting preprocessing and loss-function configurations, while its implications concern the need to balance overall classification performance, recognition of under-represented classes, and the interpretability required for expert review. The contribution should be interpreted as an evaluation of a combined workflow rather than evidence that the proposed configuration is superior: the baseline produced the highest reported accuracy in the present experiment.

Accordingly, the study addresses the following research questions:

1. How does focal loss, compared with standard cross-entropy, affect the validation loss and accuracy of an EfficientNetV2-based classifier on 22 classes of marine organisms?
2. How does AUIE preprocessing affect classification performance, and does its effect depend on the loss function used?
3. Does the combination of AUIE and focal loss direct the attention of the model, as visualized by Grad-CAM, towards the morphological features of the organisms?

METHOD

This study used a quantitative experimental approach. The overall workflow, from data preparation to evaluation and interpretation, is shown in Figure 1.

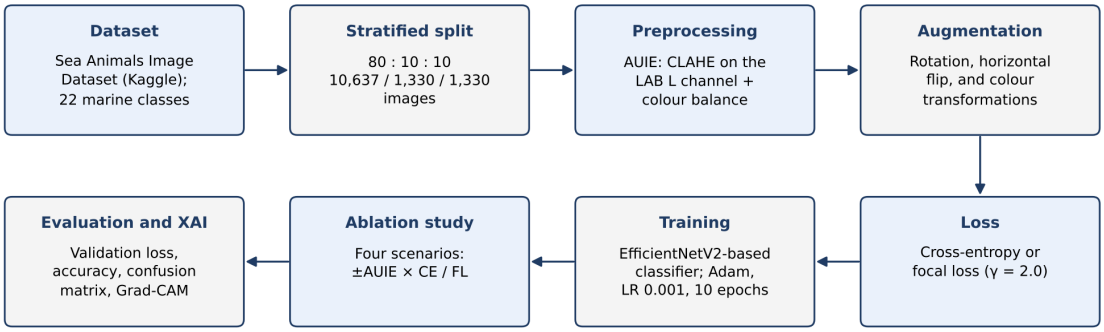


Figure 1. Experimental workflow

Research Design

The experiment followed a 2 × 2 factorial ablation design. The two independent variables were the preprocessing applied to the images (original or AUIE) and the loss function used for training (cross-entropy or focal loss). The dependent variables were validation loss and reported classification accuracy. Confusion matrices and Grad-CAM were identified as complementary qualitative analyses; however, their underlying outputs are not included in the manuscript materials available for this revision and are therefore not used as quantitative evidence in the revised interpretation. All other factors were intended to be held constant across scenarios, as summarized in Table 3. Each scenario was run once, which limits the interpretation of small differences and prevents estimation of run-to-run variability.

Table 3. Variables of the ablation design

Type of variable	Variable	Levels or setting
Independent	Image preprocessing	Original images; AUIE (CLAHE on the L channel + colour balance)
Independent	Loss function	Cross-entropy; focal loss (γ = 2.0)
Dependent	Classification performance	Validation loss; accuracy
Dependent	Model behaviour (qualitative analyses)	Confusion matrix and Grad-CAM analyses (outputs not included in this revision)
Controlled	Data and split	22 classes; stratified 80:10:10 split
Controlled	Architecture and training	EfficientNetV2-based classifier; Adam; learning rate 0.001; 10 epochs; same augmentation

Source: Data Processed

Dataset and Data Splitting

The data used in this study were the Sea Animals Image Dataset from the Kaggle platform, which consists of 22 classes of marine organisms. The dataset was divided using

a stratified split with a ratio of 80% training data (10,637 images), 10% validation data (1,330 images), and 10% test data (1,330 images), as shown in Table 4. Stratification keeps the proportion of each class approximately equal across the three subsets, so that the minority classes are represented in every partition.

The manuscript materials supplied for revision identify the dataset by name and total size but do not include its source URL, version or uploader information, the 22 class names, or class-level sample counts. These details cannot be reconstructed reliably from the aggregate partition totals and should be supplied from the original dataset record before the study is considered fully reproducible.

Table 4. Data partitions

Partition	Proportion	Number of images
Training	80%	10,637
Validation	10%	1,330
Test	10%	1,330
Total	100%	13,297

Source: Data Processed

Adaptive Underwater Image Enhancement

Image preprocessing included the application of Adaptive Underwater Image Enhancement (AUIE), which combines Contrast Limited Adaptive Histogram Equalization (CLAHE) on the lightness channel of the LAB colour space with colour balance correction to address problems of visibility and the dominance of water colour. In the LAB colour space, the L channel represents lightness, whereas the a and b channels represent the green–red and blue–yellow chromatic components. Applying contrast enhancement only to the L channel improves local contrast while limiting the distortion of hue that would occur if each RGB channel were equalized independently.

CLAHE divides the image into small tiles and equalizes the histogram of each tile separately, so that contrast is adapted to local content. To prevent the amplification of noise in homogeneous regions, the histogram of each tile is clipped at a predefined limit and the clipped counts are redistributed uniformly across the histogram before the cumulative distribution function is computed; neighbouring tiles are then combined by bilinear interpolation to avoid block artefacts. For a tile with clipped histogram h(i), M pixels, and L grey levels, the mapped intensity of a pixel with grey level v is given by Equation (1). The method follows established descriptions of adaptive histogram equalization and CLAHE (Sonali et al., 2019; Ulutas & Ustubioglu, 2021).

$$T(v) = ((L - 1) / M) \Sigma_{i=0}^{v} h(i) \quad (1)$$

Colour balance correction was applied to compensate for the colour cast of the water column, in which red light is attenuated more rapidly than blue and green light. Colour balancing that restores the relative contribution of the attenuated channels is a common

first step in underwater enhancement pipelines (Ancuti et al., 2018). In this study, AUIE was applied as a preprocessing step to the images before they were passed to the network.

The original experimental record supplied with the manuscript does not specify the CLAHE clip limit, tile-grid dimensions, or the exact colour-balance algorithm. These implementation parameters have therefore not been inferred or retrospectively assigned in this revision; they must be confirmed from the original code or experiment configuration before replication.

Data Augmentation

Spatial data augmentation (rotation and horizontal flipping) and colour augmentation were also applied to the training data. Augmentation increases the variety of training inputs and reduces overfitting, provided that the transformations preserve the label of the image (Mumuni & Mumuni, 2022). Rotation and horizontal flipping are appropriate for marine organisms because their identity does not depend on orientation, whereas colour augmentation exposes the model to variations in illumination and water colour similar to those found in real underwater scenes.The specific rotation range and colour-augmentation parameter values are not recorded in the supplied manuscript materials. Accordingly, only the augmentation types documented by the author are reported, and exact parameter values remain an author-confirmation item.

Classification Model

The classifier was built on an EfficientNetV2 backbone. EfficientNetV2 combines Fused-MBConv blocks in the early stages, where a single regular 3 × 3 convolution replaces the expansion and depthwise convolutions of the MBConv block, with MBConv blocks and squeeze-and-excitation modules in the later stages (Tan & Le, 2021). The architecture was obtained through training-aware neural architecture search and scaling, which improves both parameter efficiency and training speed compared with the original EfficientNet family.

The output of the backbone was passed to a classification layer with 22 outputs and a softmax activation, which converts the logits z_j into class probabilities as in Equation (2), where $C = 22$ is the number of classes. The supplied manuscript does not identify the EfficientNetV2 variant, whether pretrained ImageNet weights were used, the input-image resolution, or the batch size. These settings cannot be stated without access to the original training script or configuration file and should be verified by the author before final resubmission.

$$p_j = \exp(z_j) / \Sigma_{k=1}^{C} \exp(z_k) \quad (2)$$

Focal Loss

The model was trained with the loss function optimized using focal loss, which replaces standard cross-entropy in order to give more weight to minority and hard-to-classify samples. For an image whose true class has predicted probability p_t, the cross-entropy loss and the focal loss are defined in Equations (3) and (4), where $\gamma \geq 0$ is the focusing parameter (Lin et al., 2020).

$$CE(p_t) = -\log(p_t) \quad (3)$$

$$FL(p_t) = -(1 - p_t)^{\gamma} \log(p_t) \quad (4)$$

The modulating factor $(1 - pt)\gamma$ reduces the loss contribution of well-classified samples, for which pt is close to one, while leaving the loss of poorly classified samples less affected. The experiment used $\gamma = 2.0$. The supplied manuscript and experiment description do not establish whether the class-balancing weight α was implemented; therefore, its use or omission cannot be asserted in this revision. The author should verify this point against the original loss-function code. The numerical comparison of focal-loss and cross-entropy values should also be interpreted cautiously because the two loss functions have different scales. (Lin et al., 2020).

Training Configuration and Ablation Study

All scenarios were trained for 10 epochs using the Adam optimizer with a learning rate of 0.001. Evaluation was carried out through four ablation study scenarios that varied preprocessing (original images or AUIE) and the loss function (cross-entropy or focal loss), as summarized in Table 5. The supplied materials do not document a learning-rate scheduler, early-stopping rule, batch size, or training curves; none is assumed in this report. The 10-epoch schedule is retained because it is the only training duration documented in the manuscript, but it should be regarded as a limited training budget rather than an established optimum. Performance was summarized using the reported accuracy and loss values.

Table 5. Ablation study scenarios

Scenario	Preprocessing	Loss function
Baseline	Original dataset	Cross-entropy
Scenario 2	Original dataset	Focal loss ($\gamma = 2.0$)
Scenario 3	AUIE	Cross-entropy
Proposed scenario	AUIE	Focal loss ($\gamma = 2.0$)

Source: Data Processed

Accuracy is the proportion of correctly classified images among all evaluated images, as in Equation (5), where N is the number of images and $\hat{y}n$ and yn are the predicted and true labels of image n. Because accuracy can be dominated by majority classes in an imbalanced dataset, it should be complemented by per-class recall and macro-averaged F1 (Tharwat, 2021). The manuscript reports aggregate accuracy but does not identify whether the values were calculated on the validation or test partition. The revised text therefore refers to them as reported accuracy and does not describe them as test-set performance. Per-class metrics cannot be added without the corresponding predictions and labels.

$$\text{Accuracy} = (1/N) \sum_{n=1}^{N} 1[\hat{y}_n = y_n] \quad (5)$$

Explainability with Grad-CAM

The visual explanations of the models were obtained using Gradient-weighted Class Activation Mapping (Grad-CAM). Grad-CAM computes the gradient of the score y^c of class c with respect to the feature maps A^k of the last convolutional layer, averages these gradients spatially to obtain the importance weight of each channel, and forms a weighted combination of the feature maps followed by a ReLU, as in Equations (6) and (7), where Z is the number of spatial positions (Selvaraju et al., 2020). The resulting map is upsampled to the size of the input image and overlaid as a heatmap, so that regions with high values indicate the parts of the image that most increased the score of the predicted class.

$$\alpha_k^c = (1/Z) \Sigma_i \Sigma_j \partial y^c / \partial A_{ij}^k \quad (6)$$

$$L_{Grad\text{-}CAM}^c = ReLU(\Sigma_k \alpha_k^c A^k) \quad (7)$$

RESULTS AND DISCUSSION

Ablation Study Results

Based on the training logs over 10 epochs with the Adam optimizer (learning rate 0.001), the ablation study carried out on the classification network, including the baseline backbone, produced the results presented in Table 6 and Figure 2.

Table 6. Results of the ablation study

Scenario	Preprocessing	Loss	Validation loss	Reported accuracy (%)
Baseline	Original	Cross-entropy	0.6697	79.47
Scenario 2	Original	Focal loss	0.4661	79.32
Scenario 3	AUIE	Cross-entropy	0.8184	74.96
Proposed	AUIE	Focal loss	0.5503	76.32

Source: Data Processed

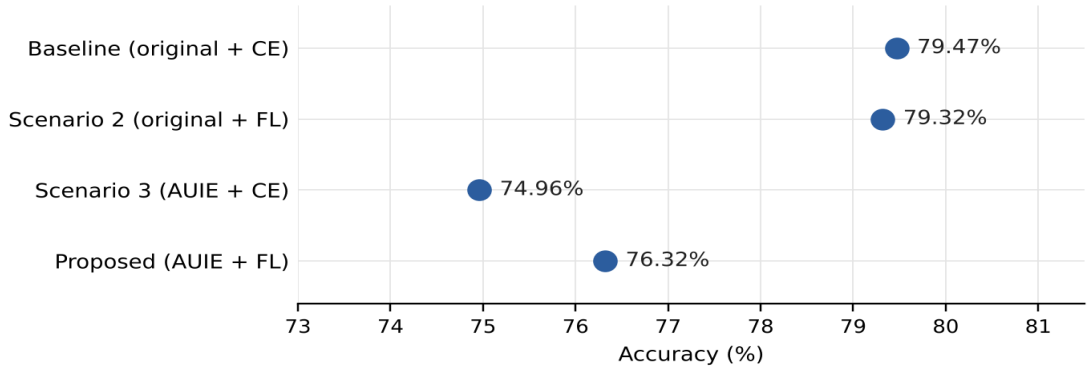


Figure 2. Accuracy of the four ablation scenarios

The baseline scenario (original dataset and cross-entropy) produced a validation loss of 0.6697 and a reported accuracy of 79.47%. Scenario 2 (original dataset and focal loss) produced a validation loss of 0.4661 and a reported accuracy of 79.32%. Scenario 3

(AUIE and cross-entropy) produced a validation loss of 0.8184 and a reported accuracy of 74.96%. The proposed scenario (AUIE and focal loss) produced a validation loss of 0.5503 and a reported accuracy of 76.32%. The baseline therefore exceeded the proposed combination by 3.15 percentage points. The manuscript does not identify the partition on which these accuracy values were calculated; they must not be interpreted as held-out test accuracy until this is verified.

Effect of Focal Loss

On the original images, replacing cross-entropy with focal loss reduced the validation loss from 0.6697 to 0.4661, while accuracy remained practically unchanged (79.47% versus 79.32%, a difference of 0.15 percentage points). The substantial decrease in loss indicates that the focal loss parameter ($\gamma = 2.0$) successfully reduced the penalty on majority samples, which are typically classified with high confidence. However, this comparison requires care: for the same predicted probabilities, focal loss is always smaller than or equal to cross-entropy, because the modulating factor $(1 - p_t)^{\gamma}$ never exceeds one.

A lower validation loss under focal loss therefore reflects the definition of the loss as well as the behaviour of the model, and loss values are strictly comparable only between scenarios that use the same loss function.With AUIE preprocessing, the reported accuracy increased from 74.96% with cross-entropy to 76.32% with focal loss, a difference of 1.36 percentage points.

This is a descriptive difference from one run per scenario and is not sufficient to establish a reliable improvement. The accompanying loss values (0.8184 and 0.5503) are not directly comparable as evidence of model improvement because the loss functions have different mathematical scales. The result is therefore limited to the observed accuracy difference in this experiment. (Lin et al., 2020).

Effect of AUIE Preprocessing

Within each loss function, AUIE reduced global accuracy: by 4.51 percentage points under cross-entropy (from 79.47% to 74.96%) and by 3.00 percentage points under focal loss (from 79.32% to 76.32%). The validation loss also increased under both loss functions. Thus, although the baseline scenario has a slightly higher global accuracy, the combination of AUIE and focal loss recovered part of the accuracy lost when AUIE was combined with cross-entropy.

Because the four scenarios form a 2 × 2 design, conditional accuracy differences can be calculated as shown in Table 7. Focal loss was associated with a −0.15 percentage-point difference on original images and a +1.36 percentage-point difference on AUIE images. AUIE was associated with a −4.51 percentage-point difference under cross-entropy and a −3.00 percentage-point difference under focal loss.

The arithmetic difference between these conditional effects is +1.51 percentage points. However, each scenario was run only once, and no paired prediction-level test or repeated-seed analysis is available. Given the approximate sampling uncertainty of

accuracy on a 1,330-image partition, these small differences cannot support a statistical interaction claim. They are reported descriptively only.

Table 7. Conditional effects of each component on accuracy

Component changed	Held constant	Accuracy change (pp)
Cross-entropy → focal loss	Original images	−0.15
Cross-entropy → focal loss	AUIE images	+1.36
Original → AUIE	Cross-entropy	−4.51
Original → AUIE	Focal loss	−3.00
Interaction (difference of conditional effects)	–	+1.51

Note. pp = percentage points; values are computed from the accuracies in Table 6.

Several factors may explain why enhancement did not increase global accuracy in this experiment. First, when a backbone is initialized with weights learned from natural images, as is common practice, enhanced images with stretched local contrast and altered colour statistics may move the input distribution further away from the pretraining distribution. Second, CLAHE can amplify noise and compression artefacts in images that already have adequate contrast, and not all images in a web-collected dataset are equally degraded.

Third, with only 10 epochs of training, the network may not have had sufficient time to adapt to the enhanced input distribution. These observations agree with the argument of Liu et al. (2020) that enhancement should be evaluated through its effect on the downstream task, not only through visual quality, and they contrast with the pipeline of Lopez-Vazquez et al. (2020), in which CLAHE-based enhancement was tailored specifically to deep-sea observatory imagery.

The absolute accuracies obtained here are lower than the accuracy reported by Kyathanahally et al. (2021) and comparable to the accuracy reported by Lopez-Vazquez et al. (2020), but such comparisons should be made with caution. The present task involves 22 visually diverse classes, the images were collected from public sources with heterogeneous quality, and training was limited to 10 epochs, whereas the related studies differ in their data, label granularity, and training protocols. The ablation design of this study is therefore more informative about the relative contribution of AUIE and focal loss than about the absolute level of performance achievable on marine images.

Minority Classes and the Confusion Matrix

The manuscript materials supplied for revision do not contain the confusion matrices or class-level prediction outputs needed to verify whether AUIE and focal loss reduced misclassification among minority classes. Accordingly, the earlier claim of reduced minority-class errors is withdrawn from the interpretation. The proposed mechanisms—improved visibility of organism contours through enhancement and greater emphasis on difficult examples through focal loss—remain plausible methodological rationales, but

they are not demonstrated as class-specific outcomes by the aggregate accuracy values reported here.

This finding also illustrates why global accuracy alone is an incomplete measure for imbalanced data. Because accuracy weights every image equally, gains on small classes can be outweighed by small losses on large classes, even when the former are more important for conservation monitoring (Langenkämper et al., 2019). For this reason, future reports of the experiment should present per-class recall, macro-averaged F1, and the confusion matrices themselves, so that the reduction in minority-class errors can be quantified and compared across scenarios (Tharwat, 2021).

Focal loss is only one of several algorithm-level strategies for imbalance. Alternatives include re-weighting the loss of each class in inverse proportion to its frequency, data-level strategies such as oversampling minority classes or undersampling majority classes, and hybrid approaches that combine both (Johnson & Khoshgoftaar, 2019). Comparative evidence in taxon identification suggests that no single strategy is uniformly best and that the ranking depends on the dataset and on whether overall accuracy or minority-class performance is emphasized (Langenkämper et al., 2019). A natural extension of this study is therefore to compare focal loss with class-weighted cross-entropy and with resampling under the same training budget. Class-balanced re-weighting based on the effective number of samples is another relevant comparison approach (Cui et al., 2019).

Interpretation with Grad-CAM

Although Grad-CAM is described in the Method section, representative heatmaps and the corresponding input images are not included in the manuscript materials supplied for revision. It is therefore not possible to verify whether the model attended to organism morphology, background regions, or other image features. The previous statement that AUIE produced more precise localization of the region of interest is consequently not retained as a finding. Grad-CAM can be discussed as the intended explainability procedure, but conclusions about localization require the actual heatmaps and, ideally, comparison with annotated organism regions.

This qualitative result is relevant to the trustworthiness of marine monitoring systems. A model whose attention falls on the organism rather than on the surrounding reef, sand, or water column is less likely to rely on spurious background correlations, which is one of the main motivations for applying XAI in computer (Linardatos et al., 2020). Nevertheless, Grad-CAM maps are coarse because they are computed at the resolution of the last convolutional layer, and visually plausible maps do not by themselves prove that an explanation is faithful to the model (Barredo Arrieta et al., 2020). A more rigorous assessment would compare heatmaps with annotated organism masks or apply quantitative tests, such as measuring the drop in confidence when the highlighted region is removed.

Implications and Limitations

The reported results show that the baseline configuration achieved the highest aggregate accuracy, while AUIE was associated with lower accuracy under both loss functions. Focal loss produced almost no accuracy difference on original images and a 1.36-percentage-point increase relative to cross-entropy on AUIE images; this single-run difference is descriptive and should not be interpreted as a confirmed improvement. Because per-class metrics and Grad-CAM outputs are unavailable, the present evidence does not establish improved minority-class recognition or enhanced interpretability.

For practical marine monitoring, the findings support evaluating preprocessing against downstream classification performance rather than assuming that visually motivated enhancement will improve accuracy. Deployment considerations also matter. Monitoring platforms such as ROVs and AUVs operate under constraints of power, memory, and communication bandwidth, so the enhancement and classification steps must be efficient enough to run on board or close to the sensor (Islam et al., 2020).

CLAHE and colour balancing are computationally light compared with deep enhancement networks, which is an advantage for such platforms, and the lightweight architectures recommended for future work would further reduce inference cost. Explanations such as Grad-CAM maps can also be used during deployment to flag predictions whose attention falls outside the organism for review by an expert. This study has several limitations. Each scenario was trained for 10 epochs and executed only once, so the stability of the observed differences across random initializations has not been assessed.

The manuscript does not identify whether reported accuracy was calculated on the validation or test partition. It also lacks per-class recall, macro-averaged F1, and the confusion-matrix outputs required to quantify minority-class performance. Although Grad-CAM is described, no heatmaps are presented for independent inspection. In addition, essential implementation details including the EfficientNetV2 variant and initialization, input resolution, batch size, CLAHE parameters, colour-balance algorithm, augmentation settings, and the use or omission of α are not available in the supplied record. The dataset's source URL, version, class names, and per-class counts are also absent. These omissions limit reproducibility and the strength of the conclusions. (Saleh et al., 2022).

Reproducibility would be strengthened by reporting the original configuration file, including the EfficientNetV2 variant and initialization, input resolution and batch size, CLAHE clip limit and tile-grid size, colour-balance algorithm, augmentation parameters, and the use or omission of focal-loss weight α. The author should also identify the partition used for each accuracy value and provide the dataset URL, version, class names, and per-class counts. Future experiments should include repeated runs with multiple random seeds, per-class recall and macro-averaged F1 on a clearly designated held-out test set, and confusion matrices for all four scenarios.

Image-quality measures such as UIQM and UCIQE may be considered in a future evaluation of enhancement quality (Panetta et al., 2016; Yang & Sowmya, 2015), while

class-balanced loss provides a relevant alternative for comparison with focal loss (Cui et al., 2019). Quantitative validation of Grad-CAM against annotated organism regions would further strengthen the explainability analysis.

CONCLUSION

This study evaluated AUIE preprocessing and focal loss in an EfficientNetV2-based classifier for 22 classes of marine organisms. The baseline scenario, using original images and cross-entropy, produced the highest reported accuracy of 79.47%, while focal loss on original images produced 79.32%, a difference of 0.15 percentage points. Applying AUIE reduced accuracy to 74.96% with cross-entropy and 76.32% with focal loss. Thus, AUIE was associated with lower global accuracy under both loss functions, whereas focal loss produced almost no accuracy difference on original images and a descriptive increase of 1.36 percentage points on enhanced images. The combined configuration did not outperform the baseline in aggregate accuracy. Because each scenario was run once, these differences should be interpreted cautiously. Further work should use repeated runs, report per-class metrics, and evaluate Grad-CAM quantitatively.

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The experiments were carried out using the author's own computational resources and the publicly available Sea Animals Image Dataset from the Kaggle platform, and no financial support was received for the implementation of this research or the publication of this article.

CONFLICT OF INTEREST

The author declares that there is no conflict of interest. This research was conducted without any commercial, financial, or personal relationship that could be construed as influencing the design of the study, the experiments with EfficientNetV2, AUIE, focal loss, and Grad-CAM, the interpretation of the results, or the decision to publish this article.

AUTHOR CONTRIBUTION

Wahyu Bontot: conceptualization, methodology, software, validation, formal analysis, investigation, data curation, visualization, writing – original draft, and writing – review and editing. As the sole author, Wahyu Bontot designed the 2 × 2 ablation study, prepared and partitioned the dataset, implemented the AUIE preprocessing, focal loss, and EfficientNetV2-based classifier, conducted the experiments and Grad-CAM analysis, interpreted the results, and approved the final version of the manuscript.

GENERATIVE ARTIFICIAL INTELLIGENCE USAGE STATEMENT

The author states that generative artificial intelligence (AI) was used in the preparation of this manuscript only to improve language clarity, readability, and grammatical structure. Generative AI was not used to design the study, generate or modify

the dataset, write or run the experimental code, produce or alter the experimental results, figures, or tables, or draw the scientific conclusions. All AI-assisted text was critically reviewed and edited by the author, and the use was carried out in compliance with applicable ethical guidelines on the use and disclosure of generative AI in scientific research. The author takes full responsibility for the accuracy, integrity, and final content of the manuscript.

REFERENCES

Ancuti, C. O., Ancuti, C., De Vleeschouwer, C., & Bekaert, P. (2018). Color Balance and Fusion for Underwater Image Enhancement. *IEEE Transactions on Image Processing*, *27*(1), 379–393. https://doi.org/10.1109/TIP.2017.2759252

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Cui, Y., Jia, M., Lin, T.-Y., Song, Y., & Belongie, S. (2019). Class-balanced loss based on effective number of samples. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9268–9277. https://doi.org/10.1109/CVPR.2019.00949

Goodwin, M., Halvorsen, K. T., Jiao, L., Knausgård, K. M., Martin, A. H., Moyano, M., Oomen, R. A., Rasmussen, J. H., Sørdalen, T. K., & Thorbjørnsen, S. H. (2022). Unlocking the potential of deep learning for marine ecology: overview, applications, and outlook. *ICES Journal of Marine Science*, *79*(2), 319–336. https://doi.org/10.1093/icesjms/fsab255

Islam, M. J., Xia, Y., & Sattar, J. (2020). Fast underwater image enhancement for improved visual perception. *IEEE Robotics and Automation Letters*, *5*(2), 3227–3234. https://doi.org/10.1109/LRA.2020.2974710

Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, *6*(1), 27. https://doi.org/10.1186/s40537-019-0192-5

Kyathanahally, S. P., Hardeman, T., Merz, E., Bulas, T., Reyes, M., Isles, P., Pomati, F., & Baity-Jesi, M. (2021). Deep learning classification of lake zooplankton. *Frontiers in Microbiology*, *12*, 746297. https://doi.org/10.3389/fmicb.2021.746297

Langenkämper, D., van Kevelaer, R., & Nattkemper, T. W. (2019). Strategies for tackling the class imbalance problem in marine image classification. *Pattern Recognition and Information Forensics. ICPR 2018. Lecture Notes in Computer Science*, *11188*, 26–36. https://doi.org/10.1007/978-3-030-05792-3_3

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollar, P. (2020). Focal Loss for Dense Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(2), 318–327. https://doi.org/10.1109/TPAMI.2018.2858826

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable AI: A review

of machine learning interpretability methods. *Entropy*, *23*(1), 18. https://doi.org/10.3390/e23010018

Liu, R., Fan, X., Zhu, M., Hou, M., & Luo, Z. (2020). Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light. *IEEE Transactions on Circuits and Systems for Video Technology*, *30*(12), 4861–4875. https://doi.org/10.1109/TCSVT.2019.2963772

Lopez-Vazquez, V., Lopez-Guede, J. M., Marini, S., Fanelli, E., Johnsen, E., & Aguzzi, J. (2020). Video image enhancement and machine learning pipeline for underwater animal detection and classification at cabled observatories. *Sensors*, *20*(3), 726. https://doi.org/10.3390/s20030726

Mumuni, A., & Mumuni, F. (2022). Data augmentation: A comprehensive survey of modern approaches. *Array*, *16*, 100258. https://doi.org/10.1016/j.array.2022.100258

Panetta, K., Gao, C., & Agaian, S. (2016). Human-visual-system-inspired underwater image quality measures. *IEEE Journal of Oceanic Engineering*, *41*(3), 541–551. https://doi.org/10.1109/JOE.2015.2469915

Raveendran, S., Patil, M. D., & Birajdar, G. K. (2021). Underwater image enhancement: a comprehensive review, recent trends, challenges and applications. *Artificial Intelligence Review*, *54*(7), 5413–5467. https://doi.org/10.1007/s10462-021-10025-z

Saleh, A., Sheaves, M., & Rahimi Azghadi, M. (2022). Computer vision and deep learning for fish classification in underwater habitats: A survey. *Fish and Fisheries*, *23*(5), 977–999. https://doi.org/10.1111/faf.12666

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, *128*(2), 336–359. https://doi.org/10.1007/s11263-019-01228-7

Sonali, Sahu, S., Singh, A. K., Ghrera, S. P., & Elhoseny, M. (2019). An approach for de-noising and contrast enhancement of retinal fundus image using CLAHE. *Optics & Laser Technology*, *110*, 87–98. https://doi.org/10.1016/j.optlastec.2018.06.061

Tan, M., & Le, Q. V. (2021). EfficientNetV2: Smaller models and faster training. *Proceedings of the 38th International Conference on Machine Learning, Proceedings of Machine Learning Research*, *139*, 10096–10106.

Tharwat, A. (2021). Classification assessment methods. *Applied Computing and Informatics*, *17*(1), 168–192. https://doi.org/10.1016/j.aci.2018.08.003

Ulutas, G., & Ustubioglu, B. (2021). Underwater image enhancement using contrast limited adaptive histogram equalization and layered difference representation. *Multimedia Tools and Applications*, *80*(10), 15067–15091. https://doi.org/10.1007/s11042-020-10426-2

Yang, M., & Sowmya, A. (2015). An underwater color image quality evaluation metric. *IEEE Transactions on Image Processing*, *24*(12), 6062–6071. https://doi.org/10.1109/TIP.2015.2491020