

ISSN 1998-7927(print) ISSN 2664-6498 (online)

DOI: <https://doi.org/10.33216/1998-7927-2026-304-6-67-72>

УДК 004.93:004.032.26

HYBRID OBJECT DETECTION IN LOW VISIBILITY UNDERWATER IMAGERY**Moroz D.V.****ГІБРИДНЕ ВИЯВЛЕННЯ ОБ'ЄКТІВ
НА ПІДВОДНИХ ЗОБРАЖЕННЯХ ЗА НИЗЬКОЇ ВИДИМОСТІ****Мороз Д.В.**

This study presents a multilevel analysis of object detection in underwater and aerial imagery captured under limited-visibility conditions. The physical causes of input-signal degradation, including wavelength-dependent absorption, scattering, non-uniform illumination, and low-contrast blur, are systematized, feature representations that can preserve object information under low contrast are considered, and a conceptual hybrid approach to robust detection is formalized. Analysis of recent research indicates that YOLO-, Faster R-CNN-, and transformer-based detectors may lose accuracy under adverse observation conditions when low-level visual features are strongly degraded, which reduces detection confidence and increases the number of false positives. To address this problem, a conceptual multi-branch architecture is proposed that coordinates detector-oriented image enhancement, physics-guided degradation estimation, attention/context mechanisms, multiscale feature fusion, and a real-time detector based on YOLO or RT-DETR. The enhancement branch is intended to be evaluated primarily by its influence on downstream localization rather than by visual-quality metrics alone. The experimental part of this study is deliberately limited to establishing a reference baseline: YOLO11n was trained and evaluated on the single-class Common Objects Underwater (COU) dataset for underwater technical-object detection. On the held-out test split, the baseline achieved Precision = 0.889, Recall = 0.832, mAP@0.50 = 0.888, and mAP@0.50:0.95 = 0.751. These values characterize only the YOLO11n baseline and do not constitute experimental validation of the complete hybrid architecture. Accordingly, the expected contributions of the enhancement, physics-guided, attention, and fusion branches are formulated as testable hypotheses to be verified by controlled ablation studies under identical data splits and training protocols. The principal contributions of the present work are the formalization of the conceptual architecture, the definition of the baseline experimental protocol, and the specification of a validation strategy for subsequent experiments.

Keywords: object detection, low visibility, underwater imagery, physics-guided learning

Introduction. Object detection requires simultaneous classification and localization of objects in an image. Formally, the output of a detector can be represented as a set

$$\mathcal{D} = (b_i, c_i, s_i)_{i=1}^N, \quad (1)$$

where b_i denotes the bounding-box parameters, c_i is the class label, and s_i is the confidence score.

High-quality detection therefore requires accurate localization, correct classification, robustness to changes in observation conditions, and acceptable real-time performance.

Modern object-detection research is dominated by two complementary architecture families. One-stage convolutional detectors of the YOLO family emphasize speed and practical deployment in real-time systems [1–4]. Transformer-based detectors, including DETR and RT-DETR, model global scene context and reduce dependence on hand-designed components such as anchors and non-maximum-suppression heuristics [5, 6]. Both families achieve high accuracy on standard benchmarks, but their performance can decline substantially when image formation differs from the training distribution.

Underwater and aerial scenes combine several independent degradation mechanisms. Underwater images are affected by wavelength-dependent absorption, forward and backscattering, non-uniform illumination, blue-green colour shifts, and loss of fine texture [7, 8]. Aerial and UAV imagery may contain fog, haze, smoke, dust, atmospheric scattering, turbulent distortion, large scale variation, and very small objects [9–11]. These factors suppress edges and local contrast,

weaken low-level CNN features, and complicate attention allocation in cluttered scenes.

A common approximation for both underwater and atmospherically degraded image formation is

$$I(x) = J(x)t(x) + A(1 - t(x)), \quad (2)$$

where $I(x)$ is the observed image, $J(x)$ is the latent clear image, A denotes ambient or background light, and $t(x)$ is the transmission map with values in the interval $[0, 1]$. For a homogeneous medium,

$$t(x) = \exp(-\beta d(x)), \quad (3)$$

where β is the attenuation coefficient and $d(x)$ is the distance to the object. Because $t(x)$ decreases with distance, small and distant targets are particularly difficult to detect in fog or turbid water [8, 10, 12].

A further difficulty is the limited availability of large, consistently annotated low-visibility detection datasets. UIEB is primarily an image-enhancement benchmark, SUIM supports semantic scene analysis, while URPC, RUOD, DUO, UTDAC2020 and related datasets address underwater detection tasks [7, 8, 13]. VisDrone and UAVDT are widely used for aerial small-object detection [9, 10]. Consequently, a strong detector alone is insufficient: preprocessing, multiscale representation, attention, and the training protocol must be coordinated with the target domain.

Unresolved issues. The unresolved methodological issue is how to combine image enhancement, physical degradation cues, attention, and a detector without improving visual appearance at the expense of localization. Existing studies often optimize enhancement and detection separately, which may amplify noise, distort object boundaries, or introduce domain-specific artefacts. This motivates a detection-friendly architecture in which all modules are evaluated with respect to the final detection objective [11, 14, 15].

Research objective and tasks. The objective of this study is to formalize a conceptual hybrid architecture for object detection in low-visibility underwater and aerial imagery and to establish a preliminary underwater YOLO11n baseline for its subsequent experimental validation. At the present stage, the experiment is intended to define the reference level; it is not used to claim an accuracy gain from the hybrid branches.

The following tasks were defined:

- 1) analyze modern object detector architectures relevant to the aerial and underwater domains;
- 2) generalize physical models of image degradation under conditions of fog, smoke, haze, and turbid water;
- 3) formalize the composition and information flow of a multi-branch hybrid framework;
- 4) establish a baseline training and evaluation protocol on a low-visibility underwater dataset;
- 5) define metrics and ablation experiments required to validate each proposed branch;

6) identify the limitations of current approaches and directions for further improvement.

Research methods. The study combines physical modelling, mathematical formalization, architectural synthesis, and a preliminary computational experiment.

1. Physical modelling of degradation. Equations (2) and (3) describe contrast attenuation and the contribution of background light. For underwater imagery, wavelength-dependent absorption is additionally considered because the red channel commonly degrades faster than the green and blue channels [7, 8].

2. Detector-oriented image enhancement. Nonlinear diffusion is used as the mathematical basis of the enhancement branch:

$$\frac{\partial U}{\partial \tau} = \nabla \cdot (c(x, \tau) \nabla U), \quad (4)$$

where U denotes the image and $c(x, \tau)$ is an edge-sensitive diffusion coefficient that suppresses noise in homogeneous regions while limiting smoothing across relevant boundaries. In the proposed framework, enhancement quality must be assessed by its effect on object localization rather than by visual quality alone [11, 15].

3. Attention mechanisms. Spatial-channel feature enhancement is represented by the scaled dot-product attention operator

$$\text{Att}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d}} \right) V, \quad (5)$$

where Q , K , and V are the query, key, and value matrices. In convolutional blocks, related functions can be implemented using CBAM-, EMA-, or GC-type modules; in transformer blocks, self-attention or efficient approximations can be applied [5, 6, 15, 16].

4. Evaluation criteria. Detection quality is assessed using IoU, Precision, Recall, F1-score, AP, and mAP. For a predicted and reference bounding box, IoU is defined as

$$\text{IoU} = \frac{|B_p \cap B_g|}{|B_p \cup B_g|}, \quad (6)$$

where B_p is the predicted bounding box and B_g is the ground-truth box. AP is the area under the precision-recall curve, while mAP averages AP over object classes. The COCO-style metric $\text{mAP}_{50:95}$ averages AP over IoU thresholds from 0.50 to 0.95 with a step of 0.05. Latency, FPS, parameter count, and FLOPs should be reported for real-time applications [5, 6].

5. Comparative architectural analysis. Detector families are compared by their feature-aggregation mechanism, ability to model global context, real-time suitability, and expected vulnerability to low-level image degradation. The comparison is methodological and is not presented as a substitute for experiments under identical conditions.

Table 1 summarizes the role of the principal detector families in a low-visibility pipeline. The table intentionally avoids direct comparison of benchmark values obtained with different image sizes, hardware platforms, and software implementations.

Table 1

Methodological comparison of detector families for low-visibility applications

Family	Mechanism	Context	Speed	Limitation
YOLO	CNN	Local / multiscale	High	Weak edges / contrast
DETR	Attention	Global	Medium	Data / compute
RT-DETR	Hybrid	Global	High	Domain shift
Hybrid framework	Multi-branch	Local / global	Variable	Latency; ablation

Conceptual hybrid framework. The proposed framework is a conceptual multi-branch network that combines an enhancement branch, a physics-guided branch, and an attention/context branch (Fig. 1). Their outputs are intended to be integrated through multiscale

feature fusion before being passed to the detector. The complete framework has not been trained end-to-end in the experiment reported below; therefore, its expected benefits are treated as testable hypotheses rather than demonstrated performance gains.

The enhancement branch increases local contrast, reduces selected photometric distortions, and produces a detector-oriented representation. The physics-guided branch estimates degradation using the transmission $t(x)$, background light A , attenuation coefficient β , and domain-specific colour and texture statistics. The attention/context branch emphasizes spatial and channel regions likely to contain informative object evidence. The branches are then combined through multiscale fusion of low-level boundaries, mid-level texture, and high-level semantic context.

The fused feature representation is expressed as

$$F_{\text{fuse}} = \phi(F_{\text{enh}}, F_{\text{phys}}, F_{\text{att}}) \quad (7)$$

where $\phi(\cdot)$ is a fusion operator implemented by concatenation followed by a 1×1 convolution, adaptive weighted summation, or attention-guided fusion. A controlled weighted form is

$$F_{\text{fuse}} = \alpha F_{\text{enh}} + \beta F_{\text{phys}} + \gamma F_{\text{att}}, \alpha + \beta + \gamma = 1. \quad (8)$$

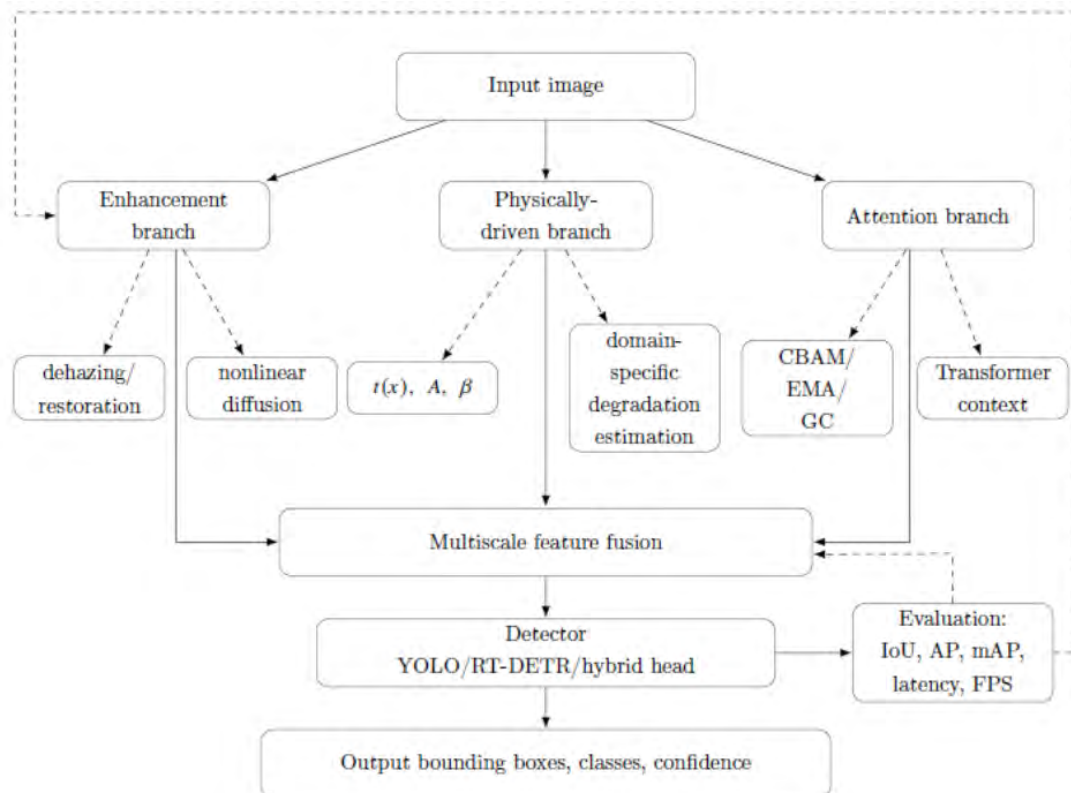


Fig. 1. Conceptual hybrid framework for low-visibility object detection

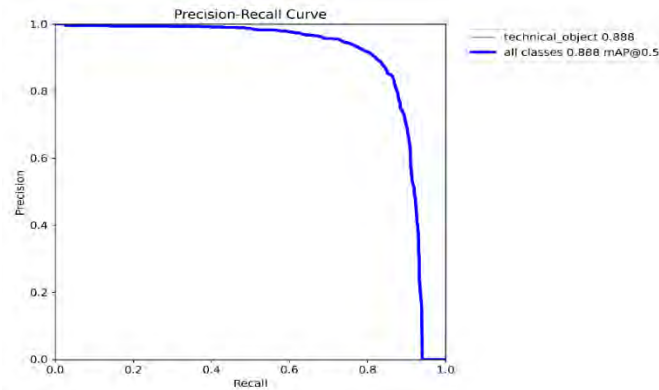


Fig. 2. PR curve for the YOLO11n on the one-class test set

The detector can be instantiated in two principal modes. A YOLO-based head prioritizes speed, multiscale detection, and deployment on resource-constrained platforms. An RT-DETR-based head provides stronger global-context modelling and end-to-end prediction without conventional NMS [6]. The appropriate configuration should be selected experimentally because the additional context of a transformer does not automatically guarantee superior robustness under domain shift.

Activation and attribution maps may be used to interpret the attention branch and to identify whether the model focuses on object regions rather than enhancement artefacts. Such visualizations support, but do not replace, quantitative ablation studies [16].

Baseline experiment (YOLO11n only). The reported computational experiment evaluates only the baseline detector and does not include the enhancement, physics-guided, attention, or fusion branches. A preliminary one-class benchmark was constructed from the COU dataset. All annotated man-made underwater objects were merged into a single category, `technical_object`, to evaluate the reference capability of detecting technical targets in degraded underwater scenes.

The data were converted to the standard YOLO detection format and divided into independent training, validation, and test subsets. The configuration file `dataset_1class.yaml` specified `nc = 1` and the unified class name. The same class definition and image geometry were retained across all subsets.

YOLO11n was selected as a lightweight baseline suitable for CPU-oriented experiments [17]. The model was trained for 10 epochs with an image size of 512 pixels and a batch size of 8. Parameters not explicitly changed were retained at the defaults of the training implementation. Diagnostic curves and intermediate statistics were generated during training. This setup should be interpreted as a preliminary reference experiment rather than a final exhaustive optimization of the detector.

Model selection was performed using validation results, after which the best checkpoint was additionally evaluated on the held-out test split. The main metrics

were Precision, Recall, $mAP@0.50$, and $mAP@0.50:0.95$. The corresponding *PR-curve* (Fig. 2) was used as the main visual representation of detection quality and as a baseline reference for future comparison with RT-DETR and hybrid low-visibility models.

The obtained YOLO11n results (Table 2) establish a quantitative reference level for one-class underwater technical-object detection on the COU benchmark. The test performance of Precision = 0.889, Recall = 0.832, $mAP@0.50 = 0.888$, and $mAP@0.50:0.95 = 0.751$ shows that the lightweight baseline is already effective on this particular split. These results should not be interpreted as evidence for the effectiveness of the proposed hybrid architecture. Instead, they define the reference against which future enhancement-only, physics-guided, attention-only, and full-fusion variants must be compared, especially under stronger visibility degradation, cross-domain shifts, blur, and low-contrast conditions.

Table 2

Baseline detection results of YOLO11n on the one-class COU

Split	Precision	Recall	$mAP@0.50$	$mAP@0.50:0.95$
Validation	0.875	0.800	0.838	0.717
Test	0.889	0.832	0.888	0.751

The normalized confusion matrix in Fig. 3 provides information beyond aggregate metrics. In the one-class setting, it shows that most labelled objects are assigned to the `technical_object` class, consistent with the reported Precision and Recall. At the same time, the background-related cells indicate that false detections and missed objects remain relevant error sources and should be analysed on the most strongly degraded images.

Because the complete hybrid framework was not trained in the reported experiment, no numerical improvement can be attributed to the enhancement, physics-guided, attention, or fusion branches. Throughout this paper, only the YOLO11n values reported in Table 2 are treated as experimentally established results; statements about the additional

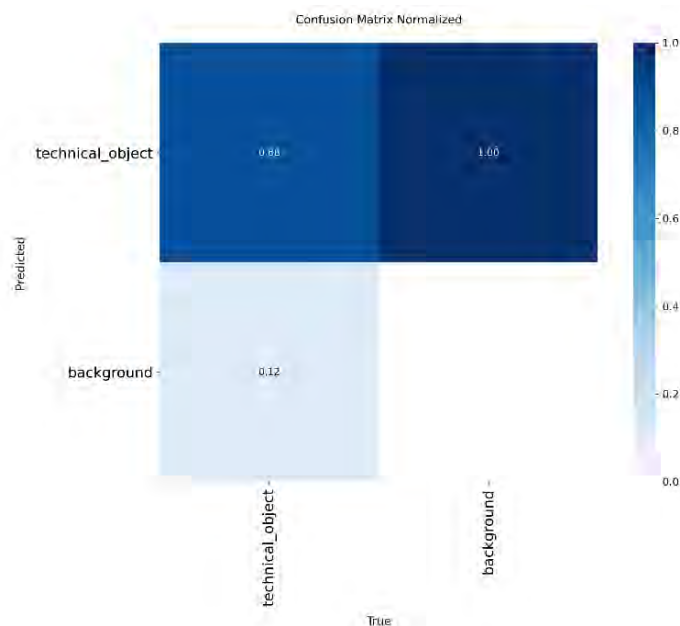


Fig. 3. Normalized confusion matrix for underwater technical object detection on the COU test split

branches describe the proposed methodology and hypotheses for subsequent ablation testing under identical train/validation/test splits.

Discussion and limitations. The analysis leads to four principal observations:

1) results reported on standard COCO-style benchmarks are useful for architecture selection but do not constitute evidence of robustness in underwater or adverse-weather imagery;

2) preprocessing must be evaluated using detection metrics because visually improved images may still distort boundaries or suppress weak objects;

3) the enhancement, physics-guided, attention, and fusion components require controlled ablation studies together with latency and complexity measurements;

4) generalization to aerial imagery remains a research hypothesis until the framework is tested on an aerial low-visibility benchmark under the same evaluation protocol.

Conclusions. This paper formalizes a conceptual hybrid framework for object detection under low-visibility conditions by coordinating detector-oriented enhancement, physical degradation cues, attention/context modelling, multiscale fusion, and a real-time detector. At this stage, this architecture is a methodological proposal: the analysis defines its information flow and the hypotheses that must be tested, but does not establish an experimental advantage over the baseline.

The experimentally confirmed part of the study is limited to the preliminary YOLO11n baseline on the one-class COU benchmark, which achieved Precision = 0.889, Recall = 0.832, mAP@0.50 = 0.888, and mAP@0.50:0.95 = 0.751 on the test split. These results provide the reference level for subsequent comparison and do not validate the complete hybrid architecture. The

next experimental stage should compare the baseline with enhancement-only, physics-guided, attention-only, and full-fusion variants on underwater datasets such as URPC, DUO, or RUOD and on an aerial benchmark such as VisDrone. Repeated runs, controlled synthetic degradations, and identical training protocols are required for statistically defensible conclusions.

The present study has three main limitations:

1) only the baseline detector, rather than the complete hybrid framework, was evaluated;

2) the experiment covered one underwater dataset and one merged class, so cross-domain and multiclass generalization were not established;

3) the reported training schedule was deliberately limited and did not include repeated runs, full hyperparameter optimization, or an ablation of computational cost.

Within these limitations, the proposed framework defines a technically coherent research direction for marine monitoring, underwater technical-object search, UAV-based inspection, and object detection in fog. Its methodological contribution is the explicit coordination of enhancement and detection within a single task-oriented pipeline; its practical effectiveness remains to be quantified in the planned ablation and cross-domain experiments.

References

1. Redmon J., Divvala S., Girshick R., Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. Pp. 779–788. DOI: <https://doi.org/10.1109/CVPR.2016.91>
2. Wang C.-Y., Bochkovskiy A., Liao H.-Y. M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. 2023 IEEE/CVF Conference

- on Computer Vision and Pattern Recognition (CVPR). 2023. Pp. 7464–7475. DOI: <https://doi.org/10.1109/CVPR52729.2023.00721>
3. Wang C.-Y., Yeh I.-H., Liao H.-Y. M. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. arXiv preprint arXiv:2402.13616. 2024. 17 p. DOI: <https://doi.org/10.48550/arXiv.2402.13616>
4. Wang A., Chen H., Liu L., Chen K., Lin Z., Han J., Ding G. YOLOv10: Real-time End-to-End Object Detection. arXiv preprint arXiv:2405.14458. 2024. 21 p. DOI: <https://doi.org/10.48550/arXiv.2405.14458>
5. Carion N., Massa F., Synnaeve G., Usunier N., Kirillov A., Zagoruyko S. End-to-End Object Detection with Transformers. Springer, 2020. Pp. 213–229. DOI: https://doi.org/10.1007/978-3-030-58452-8_13
6. Zhao Y., Lv W., Xu S., Wei X., Wang G., others. DETRs Beat YOLOs on Real-Time Object Detection. arXiv preprint arXiv:2304.08069. 2023. 14 p. DOI: <https://doi.org/10.48550/arXiv.2304.08069>
7. Li C., Guo C., Ren W., Cong R., Hou J., Kwong S., Tao D. An Underwater Image Enhancement Benchmark Dataset and Beyond. IEEE Transactions on Image Processing. 2020. Vol. 29. Pp. 4376–4389. DOI: <https://doi.org/10.1109/TIP.2019.2955241>
8. Jian M., Yang N., Tao C., Zhi H., Luo H. Underwater Object Detection and Datasets: A Survey. Intelligent Marine Technology and Systems. 2024. Vol. 2. P. 9 (23 p.). DOI: <https://doi.org/10.1007/s44295-024-00023-6>
9. Zhu P., Wen L., Du D., Bian X., Hu Q., Ling H. Vision Meets Drones: A Challenge. Springer, 2021. Pp. 444–460. DOI: https://doi.org/10.1007/978-3-030-11021-5_29
10. Tahir N. U. A., Tahir M., Ahn C. W., Hwang S. O. Object Detection in Autonomous Vehicles Under Adverse Weather: A Review of Traditional and Deep Learning Approaches. Algorithms. 2024. Vol. 17, No. 3. P. 103 (29 p.). DOI: <https://doi.org/10.3390/a17030103>
11. Kumar D., Muhammad N. Object Detection in Adverse Weather for Autonomous Driving Through Data Merging and YOLOv8. Sensors. 2023. Vol. 23, No. 20. P. 8471 (18 p.). DOI: <https://doi.org/10.3390/s23208471>
12. He K., Sun J., Tang X. Single Image Haze Removal Using Dark Channel Prior. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2011. Vol. 33, No. 12. Pp. 2341–2353. DOI: <https://doi.org/10.1109/TPAMI.2010.168>
13. Islam M. J., Edge C., Xiao Y., Luo P., Mehtaz M., Morse C., Enan S. S., Sattar J. Semantic Segmentation of Underwater Imagery: Dataset and Benchmark. 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2020. Pp. 1766–1773. DOI: <https://doi.org/10.1109/IROS45743.2020.9340821>
14. Meng X., Liu Y., Fan L., Fan J. YOLOv5s-Fog: An Improved Model Based on YOLOv5s for Object Detection in Foggy Weather Scenarios. Sensors. 2023. Vol. 23, No. 11. P. 5321 (16 p.). DOI: <https://doi.org/10.3390/s23115321>
15. Li Q., Shi H. YOLO-GE: An Attention Fusion Enhanced Underwater Object Detection Algorithm. Journal of Marine Science and Engineering. 2024. Vol. 12, No. 10. P. 1885 (22 p.). DOI: <https://doi.org/10.3390/jmse12101885>
16. Yan J., Zhou C., Zhang J., Feng H., Wang Y., Cui Y., Wang Z. Underwater Object Detection Algorithm Based on Attention Mechanism and Cross-Stage Partial Fast Spatial Pyramidal Pooling. Frontiers in Marine Science. 2022. Vol. 9. P. 1056300 (12 p.). DOI: <https://doi.org/10.3389/fmars.2022.1056300>

17. Ultralytics. Ultralytics YOLO11 Documentation. 2024. URL: <https://docs.ultralytics.com/models/yolo11/> (accessed 16 May 2026).

Мороз Д. Гібридне виявлення об'єктів на підводних зображеннях за низької видимості.

У дослідженні виконано багаторівневий аналіз задачі виявлення об'єктів на підводних та аерознімках, отриманих в умовах обмеженої видимості. Систематизовано фізичні причини деградації вхідного сигналу, зокрема залежне від довжини хвилі поглинання, розсіювання, нерівномірне освітлення та розмиття за низького контрасту; розглянуто форми подання ознак, здатні краще зберігати інформацію про об'єкти за низького контрасту, та формалізовано концептуальний гібридний підхід до надійного виявлення. Аналіз останніх досліджень свідчить, що детектори на основі YOLO, Faster R-CNN та трансформерів можуть втрачати точність за несприятливих умов спостереження, коли низькорівневі візуальні ознаки суттєво деградують, що знижує впевненість виявлення та збільшує кількість хибнопозитивних результатів. Для розв'язання цієї проблеми запропоновано концептуальну багатогілкову архітектуру, яка узгоджено поєднує орієнтоване на детектор покращення зображення, фізично обґрунтоване оцінювання деградації, механізми уваги та контексту, мультимасштабне об'єднання ознак і детектор реального часу на основі YOLO або RT-DETR. Гілку покращення передбачено оцінювати насамперед за її впливом на подальшу локалізацію об'єктів, а не лише за метриками візуальної якості. Експериментальну частину роботи свідомо обмежено формуванням еталонного базового рівня: модель YOLO11n навчено та оцінено на однокласовому наборі Common Objects Underwater (COU) для виявлення підводних технічних об'єктів. На відкладеній тестовій вибірці базова модель досягла Precision = 0.889, Recall = 0.832, mAP@0.50 = 0.888 та mAP@0.50:0.95 = 0.751. Ці значення характеризують лише базову модель YOLO11n і не є експериментальним підтвердженням повної гібридної архітектури. Відповідно, очікуваний внесок гілок покращення, фізичного оцінювання, уваги та об'єднання ознак сформульовано як гіпотези, що підлягають перевірці контрольованими абляційними експериментами за однакових розбиттів даних і протоколів навчання. Основними результатами роботи є формалізація концептуальної архітектури, визначення базового експериментального протоколу та специфікація стратегії валідації для подальших експериментів.

Ключові слова: виявлення об'єктів, низька видимість, підводні зображення, фізично кероване навчання.

Мороз Дмитро Володимирович – аспірант кафедри інженерії програмного забезпечення, Національний Університет «Одеська Політехніка».

<https://orcid.org/0009-0002-1082-5033>

e-mail: mordmytro@stud.op.edu.ua

Дата першого надходження статті 24.06.2026.

Дата прийняття статті до друку після рецензування 26.08.2026.

Дата публікації 30.09.2026.



Стаття з відкритим доступом,
відповідно до умов ліцензії
Creative Commons (CC BY 4.0)