PhysFlowNet: Learning Canonical Latent Manifolds via Spatio-Spectral Physics Priors for Underwater Object Detection

Xueting Liu[1,*], Haoyu Ji[1,*], Wenze Huang[1], Zhihao Yang[1], Yu Gao[1], Weihong Ren[1], Zhiyong Wang[1,†], and Honghai Liu[1,2]

[1] State Key Laboratory of Robotics and Systems, Harbin Institute of Technology Shenzhen, Shenzhen 518055, China
isXuetingLiu@gmail.com, wangzhiyong@hit.edu.cn
[2] Southeast University, Nanjing 210096, China

Abstract. Underwater object detection is fundamentally ill-posed due to severe light attenuation and scattering. These physical degradations induce highly non-linear geometric distortions in the visual feature space, deviating significantly from the canonical manifold of clear images and rendering standard Euclidean convolutions suboptimal. To address this, we propose PhysFlowNet, that reformulates representation rectification as a physics-guided geometric evolution on a Riemannian manifold. PhysFlowNet first extracts a joint spatial-frequency physical prior to encapsulate macroscopic scattering mechanics. Conditioned on this prior, we introduce the Parallel Physics-Residual Bottleneck (PPRB), which executes a metric-preconditioned Riemannian feature retraction to transport distorted features back to their canonical states. Moreover, to counteract cascaded smoothing during scale transitions, we introduce a Physics-Guided Manifold Downsampling (PMD) strategy to strictly preserve fine detail textures and structural boundaries. Ultimately, the network is optimized via a Unified Evidential-Contrastive Objective (UECO). Comprising a Supervised Contrastive Manifold Loss (SCML) and an Evidential Uncertainty Regulation (EUR), UECO jointly ensures latent geometric alignment and reliable uncertainty calibration. Extensive evaluations across challenging underwater benchmarks demonstrate that PhysFlowNet achieves state-of-the-art performance, establishing a principled geometric paradigm for physically degraded vision tasks.

Keywords: Underwater Object Detection · Spatio-Spectral Physical Priors · Physics-Informed Learning · Feature Rectification

1 Introduction

Underwater object detection (UOD) is critical for marine applications [12, 24, 29]. Due to wavelength-dependent absorption and particle-induced scattering

* Xueting Liu and Haoyu Ji contributed equally.
† Corresponding author.

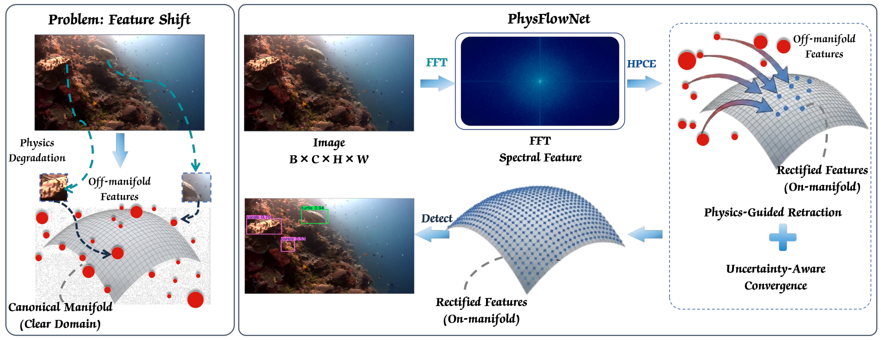

Fig. 1: Feature Rectification via Physics-Guided Riemannian Flow. Underwater degradation causes nonlinear feature drift from the clear-image manifold. PhysFlowNet uses spectral physics priors to condition a Riemannian rectified flow that geometrically realigns features, improving detection and uncertainty modeling in turbid water.

in the water medium, underwater images often exhibit contrast attenuation, color distortion, and blur. These degradation mechanisms are modeled by the Jaffe–McGlamery imaging formulation [2, 40], which explains the fundamental discrepancy between underwater imagery and standard Euclidean visual assumptions [23]. Mathematically, these degradations warp the Euclidean feature space, destroying the linear separability of semantic classes [7, 47]. Existing countermeasures face inherent limitations. Image Enhancement and Restoration (IER) [2, 20, 25, 26, 34] imposes steep computational burdens and critically tends to over-smooth crucial fine-grained details or hallucinate pseudo-structures, which misleads downstream detectors [3, 39]. Alternatively, Domain Adaptation (DA) [7,61] struggles to synthesize the depth-varying physical heterogeneity of real water columns [30]. Consequently, standard CNNs are forced into blind memorization. They consistently fail under low visibility, suffering from severe missed detections of low-contrast targets and generating high-confidence false positives driven by backscatter [12, 19, 33, 43].

To resolve this, we postulate that underwater degradation should be modeled as a physics-conditioned feature rectification process. By explicitly endowing the latent space with a Riemannian geometry, distorted features can be rigorously transported back to a canonical semantic state, as illustrated in Fig. 1.

To actualize this paradigm, we propose PhysFlowNet. Motivated by the relative stability of structural cues in the frequency domain under scattering, we extract a joint spatio–frequency prior using Hierarchical Physical Context Embedding (HPCE). Serving as a low-dimensional proxy for the degradation field, this prior dynamically conditions our Parallel Physics-Residual Bottleneck (PPRB). PPRB executes a mathematically rigorous Riemannian retraction ($x_{t+1} = \mathcal{R}_{x_t}(G^{-1/2}v)$) to neutralize physics-induced semantic shifts. Furthermore, to reduce the cascaded smoothing effects during downsampling, we introduce the Physics-Guided Manifold Downsampling (PMD) strategy to execute spatial resolution reduction. PMD explicitly preserves fine textures and structural boundaries typically attenuated by standard convolutions [11, 36, 51], guaranteeing semantic fidelity across feature pyramids. Finally, we formulate a Unified Evidential-Contrastive Objective (UECO), which combines a Supervised Contrastive Manifold Loss (SCML) for intra-class compactness on curved space and an evidential uncertainty regularization (EUR) to calibrate epistemic uncertainty and suppress backscatter-driven false positives [14].

Experiments on four underwater benchmarks (DUO, RUOD, TrashCan, and Brackish) shows that PhysFlowNet achieves state-of-the-art performance with improved efficiency, supporting explicit Riemannian-transport modeling of underwater degradation. Our contributions are: (1) We cast UOD as a physics-guided geometric rectification problem and introduce HPCE, which constructs a spatio–frequency prior embodying optical degradation cues. (2) Conditioned on this prior, PPRB performs a preconditioned Riemannian retraction to transport distorted features toward canonical representations. (3) We propose PMD to preserve fine-grained boundaries during hierarchical downsampling under scattering-induced smoothing. (4) UECO integrates SCML for manifold-topology regularization and EUR for uncertainty calibration, thereby suppressing background false positives in turbid scenes.

2 Related Work

Underwater Object Detection. UOD is challenged by wavelength-dependent absorption and backscatter, leading to contrast collapse, color shift, and spatially non-uniform veiling, while suspended particles and repetitive textures increase false positives, especially for small and low-contrast targets [50, 60]. Most UOD methods inherit terrestrial backbones, spanning two-stage (Faster R-CNN [45]) and one-stage detectors (YOLO [22,59]), and then add attention [55], multi-scale fusion [6, 8, 57], and augmentation for underwater domain shift [15, 37]. Despite progress, degradation is often treated as a black-box nuisance, leaving feature learning without explicit physics conditioning under turbidity and non-uniform illumination [58]. Consequently, data-driven adaptations can be brittle under out-of-distribution degradations, motivating physics-conditioned rectification.
Methods Guided by Physics and Geometry. The Jaffe-McGlamery model catalyzed physics-driven priors in underwater reconstruction [2, 18]. Yet, most detectors remain physics-agnostic and purely data-driven [35]. Since physical degradations inherently manifest as spectral shifts, frequency-domain learning offers a complementary alignment perspective [5,54]. However, conventional fixed amplitude swapping fails to capture spatially heterogeneous scattering. Furthermore, global domain alignment via statistical [49,56] or adversarial [13,38] methods frequently overlooks local intrinsic geometry. Transport-based rectified-flows can explicitly preserve this topology [10, 42, 53], but continuous ODE solving is computationally prohibitive for real-time detection, and unconstrained manifolds are brittle without paired anchors. Alternatively, hyperspherical embeddings enforce angular similarity [27, 31], providing an efficient geometric constraint for stable rectification. Beyond representation, softmax detectors in turbid scenes

severely suffer from over-confident errors [17,41]. While Bayesian networks quantify these uncertainties at heavy sampling costs [1,16], evidential deep learning directly models predictive uncertainty via Dirichlet evidence [14], offering an efficient mechanism to suppress scattering-induced false positives. Motivated by these threads, our method integrates physics-guided Spatio-Spectral conditioning and geometric rectification, and evidential uncertainty regulation to improve detection robustness under heterogeneous underwater degradation.

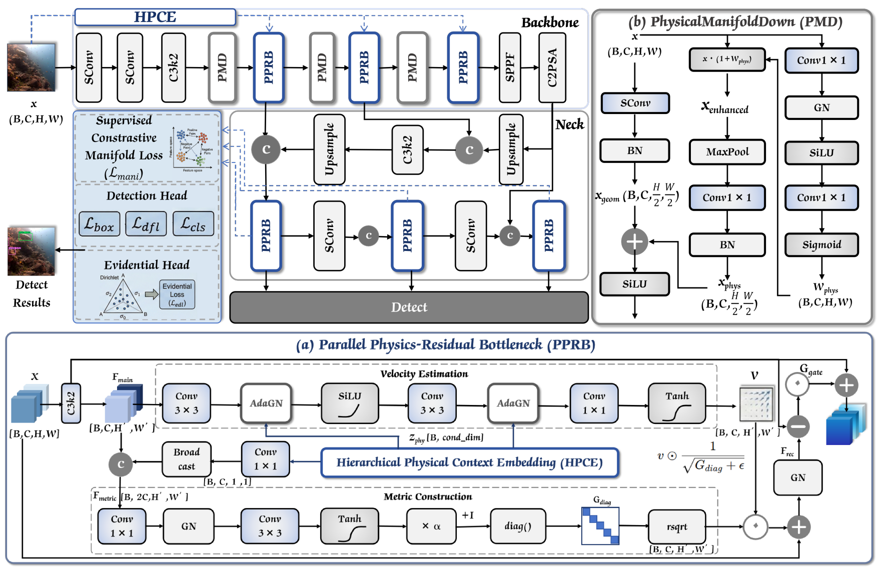

Fig. 2: Overall architecture of PhysFlowNet. The PPRB corrects semantic shifts through a dual-stream Riemannian feature retraction. This retraction process is dynamically modulated by a unified physical condition z_{phy}, which is extracted by the HPCE. During spatial scale transitions, the PMD module is employed to preserve detail textures and boundaries. Finally, the entire network is optimized via a unified evidential-contrastive objective for geometric alignment and uncertainty regulation.

3 Methodology

In this section, we present PhysFlowNet, which casts underwater object detection as a physics-conditioned manifold rectification problem. The framework first extracts a unified physical prior via Hierarchical Physical Context Embedding (HPCE). Conditioned on this prior, Parallel Physics-Residual Bottleneck (PPRB) performs a metric-preconditioned retraction for feature transport, while Physics-Guided Manifold Downsampling (PMD) preserves boundary cues across scale transitions. Training is further regularized by a Unified Evidential-

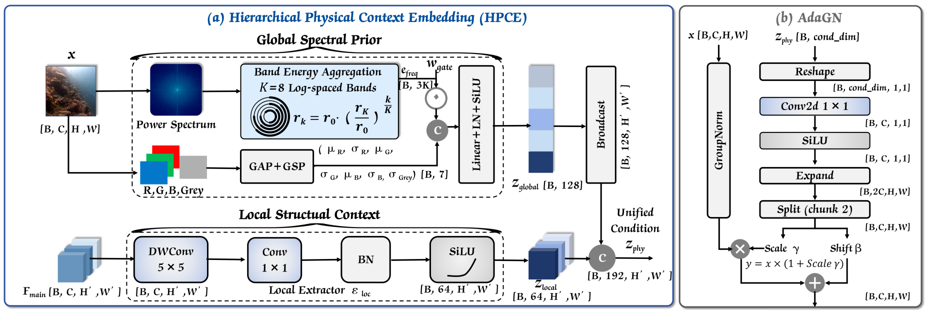

Fig. 3: Architecture of the Hierarchical Physical Context Embedding (HPCE) module. HPCE formulates a unified physical condition z_{phy} via two parallel branches. The global branch captures macroscopic degradation statistics by integrating frequency-domain power spectra and spatial-domain metrics. Concurrently, the local branch extracts spatially variant structural cues from the main feature stream F_{main}.

Contrastive Objective (UECO), combining a supervised contrastive manifold loss with evidential uncertainty regulation. The architecture is illustrated in Fig. 2.

3.1 Problem Formulation: A Riemannian Flow Perspective

Underwater image degradation, as commonly formalized by the Jaffe–McGlamery model (Eq. 1), distorts the canonical feature manifold into a heterogeneous degraded space $\mathcal{M}_d$.

$$I(x) = J(x)e^{-\eta d(x)} + (J * S)(x) + B_\infty \left(1 - e^{-\eta d(x)}\right), \tag{1}$$

where $J(x)$ denotes the clean radiance, η is the attenuation coefficient, $d(x)$ signifies the depth, S is the Point Spread Function (PSF) inducing blur, $*$ denotes the convolution operator, and B_∞ encapsulates the veiling light driven by backscattering. To restore feature separability, we reframe the representation rectification as a Riemannian flow. Given an initial degraded feature $\mathbf{h} \in \mathcal{M}_d$ and physics priors z_{phy}, we apply a retraction update along the geodesic:

$$\mathbf{h}_{rec} = \mathcal{R}_{\mathbf{h}}\Big(G(\mathbf{h}, z_{phy})^{-\frac{1}{2}} \cdot \mathcal{V}(\mathbf{h}, z_{phy})\Big), \tag{2}$$

where $\mathcal{V}$ is the velocity vector field, and the metric tensor G serves as an adaptive preconditioner correcting spatial and spectral distortions. Complete derivations and implementation details are provided in the Appendix.

3.2 Hierarchical Physical Context Embedding (HPCE)

To precisely modulate the velocity field $\mathcal{V}$ governing the rectification trajectory, the physical condition z_{phy} must encapsulate both global degradation statistics (e.g., water turbidity, color shifts) and local structural semantics (e.g., uneven illumination, localized contrast variations). To this end, we propose the Hierarchical Physical Context Embedding (HPCE) (Fig. 3 (a)), which comprises global and local branches to formulate a unified conditional representation.

1) Global Spectral and Statistical Prior. We directly extract frequency- and spatial-domain statistics from the input image to construct a joint physical proxy. Formally, given an input image $\mathbf{x} \in \mathbb{R}^{B\times C\times H\times W}$, we compute its power spectrum via the 2D Real Fast Fourier Transform (rFFT):

$$\mathcal{P}(\mathbf{x})(u,v) = |\mathcal{F}(\mathbf{x})(u,v)|^2. \tag{3}$$

For each band B_k, we aggregate the band energy using a learnable kernel W_k that is normalized within the band by Softmax:

$$\tilde{W}_k(u,v) = \mathrm{Softmax}\big(W_k(u,v)\big), \qquad \sum_{(u,v)\in B_k} \tilde{W}_k(u,v) = 1. \tag{4}$$

The channel-wise band descriptor is then defined as

$$e_{k,c} = \log\left(\sum_{(u,v)\in B_k} \tilde{W}_k(u,v)\,\mathcal{P}_c(\mathbf{X})(u,v) + \varepsilon\right), \qquad k = 1,\dots,K. \tag{5}$$

Concatenating all channels yields the spectral prior:

$$\mathbf{e}_{\mathrm{freq}} = \mathrm{Concat}\big([e_{1,R},\dots,e_{K,R}],\ [e_{1,G},\dots,e_{K,G}],\ [e_{1,B},\dots,e_{K,B}]\big) \in \mathbb{R}^{3K}. \tag{6}$$

Complementary to the frequency representation, we extract the channel-wise mean μ_c and standard deviation σ_c ($c \in \{R,G,B\}$) from the RGB image X to capture color shifts and contrast degradation driven by wavelength-dependent absorption and particle scattering, respectively. By concatenating these metrics with the global grayscale standard deviation σ_{gray}, which reflects the overarching contrast degradation—we formulate the spatial statistical prior $\mathbf{s}_{spatial} \in \mathbb{R}^7$.

To bolster numerical stability and adaptively recalibrate the significance of frequency bands, we introduce a learnable band-gating vector $\mathbf{w}_{gate} \in \mathbb{R}^K$ to reweight $\mathbf{e}_{freq}$. Then the global physical descriptor $z_{global} \in \mathbb{R}^{128}$ is generated:

$$z_{\mathrm{global}} = \mathrm{SiLU}\Big(\mathrm{LN}\big(\mathbf{W}_1\,\mathrm{Concat}\big(\mathbf{w}_{\mathrm{gate}} \odot \mathbf{e}_{\mathrm{freq}},\, \mathbf{s}_{\mathrm{spatial}}\big) + \mathbf{b}_1\big)\Big). \tag{7}$$

Here, $\mathbf{W}_1$ and $\mathbf{b}_1$ are the learnable parameters of the linear layer.

2) Local Structural Context. To capture spatially variant degradation, we process the main feature map F_{main} extracted from the preceding C3k2 block via convolutions and activation to get $z_{local} \in \mathbb{R}^{B\times 64\times H'\times W'}$:

$$z_{local} = \mathrm{SiLU}(\mathrm{BN}(\mathrm{Conv}_{1\times1}(\mathrm{DWConv}_{5\times5}(F_{main})))) \,. \tag{8}$$

Unified Conditional Representation. Finally, z_{global} is broadcast and concatenated with z_{local} to form the unified physical condition $z_{\mathrm{phy}} \in \mathbb{R}^{B\times 192\times H'\times W'}$. z_{phy} acts as the physical conditioning signal for the PPRB, guiding both feature rectification dynamics and their associated geometric regularization.

3.3 Parallel Physics-Residual Bottleneck (PPRB)

To operationalize the rectification step defined in Sec. 3.1, we introduce the PPRB (Fig. 2) (a). PPRB follows a dual-stream design in which a standard semantic branch preserves the backbone representation, while a physics-conditioned rectification branch explicitly constructs a bounded velocity field $\mathbf{v}$ and a diagonal SPD metric G_{diag} to realize a preconditioned Riemannian retraction.

The Rectification Stream: Riemannian Flow To approximate the exact analytical transport delineated in Eq. (2), the Rectification Stream is architecturally bifurcated into two foundational sub-modules: Velocity Estimation and Metric Construction.

Velocity Estimation ($\mathcal{V}$). This sub-module computes the explicit update direction $\mathbf{v} \in \mathbb{R}^{B\times C\times H\times W}$ along the manifold tangent space. To ensure the velocity field reactively adapts to varying environmental degradation intensities, we first refine the semantic feature by a projection and then modulate the main semantic features using the prior z_{phy} via Adaptive Group Normalization (AdaGN):

$$\tilde{F} = \text{Conv}_{3\times 3}(F_{\text{main}}), (\boldsymbol{\gamma}_z, \boldsymbol{\beta}_z) = \Psi(z_{\text{phy}}), \tag{9}$$

$$\text{AdaGN}(\tilde{F}, z_{\text{phy}}) = \text{GN}(\tilde{F}) \odot (\mathbf{1} + \boldsymbol{\gamma}_z) + \boldsymbol{\beta}_z, \tag{10}$$

where $\boldsymbol{\gamma}_z, \boldsymbol{\beta}_z \in \mathbb{R}^{B\times C\times 1\times 1}$ are generated through a linear projection of z_{phy}, $\Psi : \mathbb{R}^{B\times d} \to \mathbb{R}^{B\times C\times 1\times 1}$ predicts the scale–shift parameters which are broadcast to $H' \times W'$. The velocity field is formulated through a physics-guided residual mapping, encompassing the AdaGN layer, a non-linear activation (e.g., SiLU), and a convolutional projection:

$$\tilde{F}_1 = \text{Conv}_{3\times 3}\Big(\sigma\big(\text{AdaGN}(\tilde{F}, z_{\text{phy}})\big)\Big), \tag{11}$$

$$\mathbf{v} = \tanh\Big(\text{Conv}_{3\times 3}\big(\text{AdaGN}(\tilde{F}_1, z_{\text{phy}})\big)\Big). \tag{12}$$

The terminal $\tanh(\cdot)$ activation explicitly enforces a bounded constraint on the velocity amplitude, effectively suppressing erratic, large-step gradient updates during early optimization.

Metric Construction (G_{diag}). We introduce the Diagonal Symmetric Positive Definite (SPD) Metric Estimator to explicitly computes the Riemannian preconditioning term $(G+\varepsilon)^{-\frac{1}{2}}$ defined in Eq. (2) by approximating the local geometry via a channel-wise diagonal tensor $G_{\text{diag}} = \text{diag}(g_1, \ldots, g_C) \in \mathbb{R}^{B\times C\times H'\times W'}$. The physical prior z_{phy} is first projected via convolution and spatially broadcasted to align with the resolution of the main semantic features F_{main}. It is then concatenated with F_{main} to form a joint embedding $F_{\text{metric}} \in \mathbb{R}^{B\times 2C\times H'\times W'}$:

$$F_{\text{metric}} = \text{Concat}\Big(F_{\text{main}}, \text{Broadcast}\big(\text{Conv}_{1\times 1}(z_{\text{phy}})\big)\Big). \tag{13}$$

F_{metric} is processed by a dedicated bottleneck module to predict the spatial curvature variations. The module applies a sequential cascade of a dimension-reducing 1×1 convolution, Group Normalization (GN), and a spatial 3×3

convolution. The intermediate output is then bounded and shifted to formulate the diagonal elements:

$$\mathbf{g} = \mathrm{Conv}_{3\times3}\Big(\mathrm{GN}\big(\mathrm{Conv}_{1\times1}(F_{\mathrm{metric}})\big)\Big), \tag{14}$$

$$G_{\mathrm{diag}} = \mathrm{diag}\Big(\mathbf{1} + \alpha \cdot \tanh(\mathbf{g})\Big), \tag{15}$$

where $\mathrm{diag}(\cdot)$ constructs the explicit diagonal tensor field at each spatial coordinate (x, y), $\mathbf{1}$ represents the all-ones vector, and the scalar $\alpha = 0.9$ bounds the maximum allowable geometric distortion. We zero-initialize exclusively the terminal 3×3 convolutional weights, reliably anchoring the initial spatial metric field to the identity matrix $\mathbf{I}$ (since $\tanh(\mathbf{0}) = \mathbf{0}$). This initialization guarantees that the feature manifold degenerates into a flat Euclidean space uniformly across all spatial locations at the onset of training. As training progresses, the physical guidance continuously calibrates the diagonal elements, executing a pixel-level adaptive stretching or compression along specific tangent space dimensions.

Preconditioned Retraction. Given $\mathbf{v}$ and G_{diag}, we perform a metric-precondit ioned retraction:

$$F_{\mathrm{rec}} = \mathrm{GN}\Big(F_{\mathrm{main}} + \mathbf{v} \odot (G_{\mathrm{diag}} + \varepsilon)^{-\frac{1}{2}}\Big), \tag{16}$$

where ε is a small constant. Finally, we inject only the corrective displacement using a learnable spatial gate and a channel-wise scale:

$$G_{\mathrm{gate}} = \sigma\Big(\mathrm{Conv}_{1\times1}\big(\mathrm{Concat}(F_{\mathrm{main}}, F_{\mathrm{rec}})\big)\Big) \in \mathbb{R}^{B\times C\times H'\times W'}, \tag{17}$$

$$F_{\mathrm{out}} = F_{\mathrm{main}} + \boldsymbol{\gamma} \odot G_{\mathrm{gate}} \odot (F_{\mathrm{rec}} - F_{\mathrm{main}}), \tag{18}$$

where $\boldsymbol{\gamma} \in \mathbb{R}^{1\times C\times 1\times 1}$ is initialized to zero. This initialization preserves the original backbone behavior at the beginning of training and enables a progressive activation of the rectification stream.

3.4 Physics-Guided Manifold Downsampling (PMD)

Standard strided convolutions downsample by aggregating local neighborhoods and thus introduce an implicit low-pass effect. Underwater scattering already behaves as a strong optical low-pass filter, so naïvely applying strided downsampling further attenuates weak boundary cues and reduces the separability of small objects. From a physical–geometric viewpoint, this scale transition should preserve two complementary aspects: the *macro* spatial layout that supports coarse semantics and the *micro* locally prominent responses (e.g., boundary and corner activations) that are most vulnerable to compounded smoothing. Motivated by this, we propose the Physics-Guided Manifold Downsampling (PMD) module (Fig. 2(b)), which implements a structure-preserving scale transition through two parallel pathways.

Geometric Path. We employ a strided convolution to perform the primary resolution reduction and obtain the baseline geometric representation y_{geom}.

Physical Prior Path. To counteract the cascaded low-pass effect induced by scattering and downsampling, we introduce a lightweight predictor $\mathcal{E}_{\text{phy}}(\cdot)$ to estimate a physics-guided modulation and apply it in a single mapping:

$$y_{\text{phy}} = \text{Conv}_{1\times 1}\Big(\text{MaxPool}\Big(x \odot \Big(1 + \sigma\big(\mathcal{E}_{\text{phy}}(x)\big)\Big)\Big)\Big), \tag{19}$$

where $\sigma(\cdot)$ denotes the sigmoid function. This design applies residual gating prior to MaxPooling, yielding a degradation-aware reweighting and a peak-preserving downsampling step. Unlike linear aggregation that averages corrupted neighbors, MaxPooling propagates locally salient responses, thereby retaining fine textures and boundary/corner cues that survive scattering and would otherwise be diluted during subsequent scale transitions. In this way, the rectified structural details produced by the upstream manifold transport are better preserved across the feature pyramid. Finally, the two pathways are aggregated by element-wise summation followed by a non-linear activation:

$$y = \phi\big(y_{\text{geom}} + y_{\text{phy}}\big), \tag{20}$$

where $\phi(\cdot)$ denotes the activation function. Overall, PMD enables effective scale reduction while better preserving boundary contrast and local discriminability under underwater degradation.

3.5 Unified Evidential-Contrastive Objective (UECO)

Under severe turbidity and scattering, rectified features may still exhibit residual ambiguity and miscalibration driven by background-dominated statistics. We therefore couple a supervised contrastive manifold regularizer with an evidential uncertainty term to jointly enforce separability and calibration.

Supervised Contrastive Manifold Loss (SCML). We regularize multi-level head features with a supervised contrastive loss on assigner-selected foreground anchors, avoiding heuristic GT-center sampling. Given $\{F^{(\ell)}\}_{\ell=1}^{L}$ with $F^{(\ell)} \in \mathbb{R}^{B\times C\times H_\ell \times W_\ell}$ and positive indices $\mathcal{I}^{(\ell)}$, we extract $\mathbf{f}_i \in \mathbb{R}^C$ for each $i \in \mathcal{I}^{(\ell)}$ and normalize $\mathbf{z}_i = \mathbf{f}_i/\|\mathbf{f}_i\|_2 \in \mathbb{S}^{C-1}$. Define $\mathcal{P}(i) = \{p \in \mathcal{I}^{(\ell)} \mid y_p = y_i,\ p \neq i\}$, $\mathcal{A}(i) = \mathcal{I}^{(\ell)} \setminus \{i\}$, and $\mathcal{M}^{(\ell)} = \{i \in \mathcal{I}^{(\ell)} \mid |\mathcal{P}(i)| > 0\}$; the per-level loss is

$$\mathcal{L}^{(\ell)}_{\text{supcon}} = -\frac{1}{|\mathcal{M}^{(\ell)}|} \sum_{i\in\mathcal{M}^{(\ell)}} \frac{1}{|\mathcal{P}(i)|} \sum_{p\in\mathcal{P}(i)} \log \frac{\exp(\mathbf{z}_i^\top \mathbf{z}_p/\tau)}{\sum_{a\in\mathcal{A}(i)} \exp(\mathbf{z}_i^\top \mathbf{z}_a/\tau)}, \tag{21}$$

where τ is the temperature. We apply class-balanced subsampling (e.g., cap $K = 16$ anchors per class) to reduce quadratic cost. The multi-level regularizer is

$$\mathcal{L}_{\text{mani}} = \frac{1}{|\mathcal{L}_{\text{valid}}|} \sum_{\ell\in\mathcal{L}_{\text{valid}}} \mathcal{L}^{(\ell)}_{\text{supcon}}, \qquad \mathcal{L}_{\text{valid}} = \{\ell \mid |\mathcal{M}^{(\ell)}| \geq 1\}. \tag{22}$$

Evidential Uncertainty Regulation (EUR). On the same foreground anchors, we adopt EDL to obtain Dirichlet parameters from logits $\mathbf{s}_i \in \mathbb{R}^C$ via clipped Softplus evidence:

$$\hat{s}_{i,c} = \mathrm{clip}(s_{i,c}, -\kappa, \kappa), \quad e_{i,c} = \mathrm{softplus}(\hat{s}_{i,c}), \quad \alpha_{i,c} = e_{i,c} + 1, \quad S_i = \sum_{c=1}^{C} \alpha_{i,c}. \tag{23}$$

Let $\mathbf{y}_i \in [0,1]^C$ be the assigner-provided soft target. We stop its gradient and normalize it as

$$\tilde{y}_{i,c} = \frac{y_{i,c}}{\sum_{c=1}^{C} y_{i,c} + \varepsilon}, \qquad \mathbf{y}_i \leftarrow \mathrm{stopgrad}(\mathbf{y}_i). \tag{24}$$

Using the digamma function $\psi(\cdot)$, the evidential risk is

$$\mathcal{L}_{\mathrm{risk}}^{(i)} = \sum_{c=1}^{C} \tilde{y}_{i,c} \Big(\psi(S_i) - \psi(\alpha_{i,c}) \Big). \tag{25}$$

We regularize mainly the non-target mass via

$$\tilde{\alpha}_{i,c} = (\alpha_{i,c} - 1)(1 - \tilde{y}_{i,c}) + 1. \tag{26}$$

Let $\mathcal{I} = \bigcup_{\ell:\, |\mathcal{I}^{(\ell)}|>0} \mathcal{I}^{(\ell)}$; the evidential loss is

$$\mathcal{L}_{\mathrm{edl}} = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \Big[\mathcal{L}_{\mathrm{risk}}^{(i)} + \lambda_t \cdot \mathrm{KL}\big(\mathrm{Dir}(\tilde{\boldsymbol{\alpha}}_i) \,\|\, \mathrm{Dir}(\mathbf{1})\big) \Big], \tag{27}$$

where the annealing factor $\lambda_t = \min(1,\, t/T_{\mathrm{anneal}})$ stabilizes early training at epoch t.

Total Objective Function. The overall objective combines standard detection losses with the proposed manifold and evidential regularizers:

$$\mathcal{L}_{\mathrm{total}} = \lambda_1 \mathcal{L}_{\mathrm{box}} + \lambda_2 \mathcal{L}_{\mathrm{dfl}} + \lambda_3 \Big(\mathcal{L}_{\mathrm{bce}} + \lambda_4 \mathcal{L}_{\mathrm{edl}} \Big) + \lambda_5 \mathcal{L}_{\mathrm{mani}}, \tag{28}$$

where $\mathcal{L}_{\mathrm{box}}$, $\mathcal{L}_{\mathrm{dfl}}$, and $\mathcal{L}_{\mathrm{bce}}$ denote the box regression loss, distribution focal loss, and binary cross-entropy, respectively. The weights $\lambda_{(\cdot)}$ balance gradient contributions from localization, classification, and the proposed regularizers.

4 Experiment

4.1 Setting

We evaluate PhysFlowNet on four underwater benchmarks, including DUO [33], Brackish [43], TrashCan [19] and RUOD [12]. Performance is reported using COCO-style AP (AP@[.50:.95]) together with AP_{50} and AP_{75}. All models are trained from scratch for 300 epochs on a single NVIDIA RTX 4090 GPU with 640×640 inputs and batch size 16 using SGD (lr= 0.01, momentum= 0.937, weight decay= 5×10^{-4}) and a learning-rate scheduler.

4.2 Comparisons with the SOTA

Tab. 1 shows PhysFlowNet achieves a new SOTA on DUO. It uses Riemannian manifold rectification to handle scattering, yielding a +8.5% AP gain over YOLO11 [22] and beating WFDA [36] by +1.0% AP_{75}. Operating at 14.0 GFLOPs and 39.7 FPS, it demonstrates a vastly superior efficiency-accuracy trade-off compared to dense spatial attention. As shown in Tab. 2, on the TrashCan dataset, PhysFlowNet achieves 4.1% improvement over YOLO11. It also maintains top-tier performance on RUOD (86.4% AP_{50}) and Brackish (86.1% AP). This steady performance confirms that our identity-initialized Riemannian metric adapts its curvature to local degradation without overfitting to a specific dataset bias.

Table 1: Performance comparison on DUO. GOD: General Object Detection and UOD: Underwater Object Detection. Echi—Echinus; Holo—Holothurian; Scal—Scallop; Star—Starfish.

	Methods	**AP↑**	**AP$_{50}$↑**	**AP$_{75}$↑**	**Echi**	**Holo**	**Scal**	**Star**	**Params (M)↓**	**FLOPs (G)↓**	**FPS↑**
GOD	RetinaNet [32]	57.8	78.9	63.0	79.4	68.2	51.1	74.9	55.38	83.2	16.2
	Faster R-CNN [45]	61.2	82.7	69.6	70.4	61.4	41.9	71.4	41.3	120	13.5
	Cascade R-CNN [4]	61.6	80.5	72.4	69.0	61.9	41.9	72.0	88.15	140	10.3
	DetectoRS [44]	60.8	81.4	69.3	69.5	60.9	41.1	70.2	123.23	90.03	6.9
	YOLOv7 [48]	49.8	85.4	64.2	73.7	66.3	50.8	74.5	6.2	13.8	40.1
	YOLOv8 [21]	59.1	88.1	72.8	75.2	64.8	54.7	70.9	3.2	8.7	51.2
	YOLO11 [22]	60.8	89.2	73.9	75.5	64.3	55.2	73.5	**2.6**	**6.7**	**55.7**
UOD	Boosting R-CNN [46]	63.7	79.0	72.3	70.0	64.3	46.6	74.8	45.95	169	22.0
	RoIAttn [28]	62.2	82.8	69.5	70.7	62.2	40.5	71.4	55.23	331.7	9.7
	ERL-Net [9]	63.7	81.9	72.2	70.8	66.6	45.4	73.7	45.95	54.8	8.9
	GCC-Net [8]	69.1	87.8	76.3	75.2	68.2	56.3	76.7	36.74	300.79	15.4
	AMSP-UOD [59]	40.1	78.5	–	**87.5**	60.6	42.5	77.5	10.4	23.8	34.5
	WFDA [36]	66.3	90.4	79.4	82.3	70.9	60.0	80.0	2.6	7.2	54.0
	PhysFlowNet (Ours)	**69.3**	**92.7**	**80.4**	91.9	**70.7**	**64.9**	**85.9**	8.19	14.0	39.7

4.3 Ablation Study

We conduct an incremental ablation on TrashCan under an identical training protocol to isolate the contribution of each component (Table 3), while explicitly separating the physical prior design from the rectification mechanism. The baseline (YOLO11) achieves 61.9/45.1 AP_{50}/AP. Adding PMD improves performance to 63.0/46.4, indicating that physics-guided downsampling alleviates scattering-amplified information loss during scale reduction. Incorporating HPCE further raises the score to 63.3/46.8, showing that the learned physical context provides additional discriminative conditioning. To disentangle architectural effects from geometric ones, we next enable PPRB with an Euclidean update ($G = \mathbf{I}$), obtaining 63.6/47.2, which suggests that dual-stream decoupling with AdaGN yields limited but consistent gains. In contrast, activating

Table 2: Benchmarking results between PhysFlowNet and other SOTA methods on Brackish, TrashCan, and RUOD.

Methods	Brackish		TrashCan		RUOD	
	AP↑	AP_{50}↑	AP↑	AP_{50}↑	AP↑	AP_{50}↑
RetinaNet [32]	74.0	91.3	29.4	53.8	48.2	74.3
Faster R-CNN [45]	61.2	62.9	31.2	55.3	54.3	74.9
Cascade R-CNN [4]	80.7	98.5	33.6	52.7	56.0	79.7
YOLOv7 [48]	71.3	90.5	26.1	48.8	47.2	76.4
YOLOv8 [21]	80.1	97.7	44.2	61.5	58.0	82.1
YOLO11 [22]	80.4	98.1	45.1	61.9	58.1	81.9
Boosting R-CNN [46]	79.6	97.4	36.8	57.6	52.3	79.8
RoIAttn [28]	79.3	91.2	32.5	56.8	49.7	75.2
ERL-Net [9]	85.4	<u>98.8</u>	37.0	58.9	53.2	80.5
GCC-Net [8]	80.5	98.3	41.3	61.2	56.1	83.2
AMSP-UOD [59]	–	–	–	–	**65.2**	<u>86.1</u>
WFDA [36]	82.6	98.5	46.9	63.9	63.4	<u>86.1</u>
SFUDNet [52]	<u>86.0</u>	98.6	–	–	–	–
PhysFlowNet (Ours)	**86.1**	**99.1**	**49.2**	**65.4**	<u>64.3</u>	**86.4**

Table 3: Step-wise ablation study on the TrashCan dataset. Euc.: Euclidean residual projection; Riem.: our proposed Riemannian retraction flow.

PMD	HPCE	PPRB	Geometry	$\mathcal{L}_{mani}$ / $\mathcal{L}_{edl}$	AP_{50}	AP
			Euc. (Base)	/	61.9	45.1
✓			Euc.	/	63.0	46.4
✓	✓		Euc.	/	63.3	46.8
✓	✓	✓	Euc.	/	63.6	47.2
✓	✓	✓	Riem.	/	**64.9**	**48.4**
✓	✓	✓	Riem.	✓/	<u>65.1</u>	<u>48.8</u>
✓	✓	✓	Riem.	✓/ ✓	**65.4**	**49.2**

the diagonal metric and the metric-preconditioned retraction (G_{diag}) produces a clear jump to 64.9/48.4, highlighting the importance of metric-guided updates. Finally, adding $\mathcal{L}_{\text{mani}}$ and $\mathcal{L}_{\text{edl}}$ further reaches 65.4/49.2, completing the full model. More detailed ablations for each module are provided in the Appendix.

4.4 Qualitative Analysis

Visualization of Feature Manifolds. We assess representation quality via t-SNE on RoI features from 800 randomly sampled instances in the DUO and RUOD validation sets. RoI embeddings are extracted from the P4 layer before the detection head. To avoid manual domain labeling, we cluster the 128-D HPCE physical-prior embeddings using K-Means into three degradation domains, which are encoded by marker shapes, while semantic classes are encoded by colors. As shown in Fig. 4(a,b), the baseline exhibits strong entanglement: samples of the same class split into domain-specific sub-clusters. In contrast, PhysFlowNet yields more compact class-wise clusters with improved inter-class separation and reduced domain stratification, indicating decreased sensitivity to heterogeneous underwater degradations.

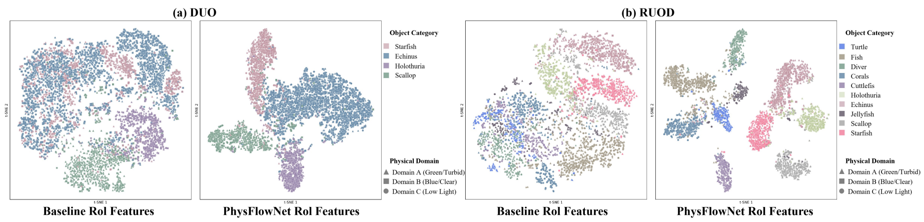

Fig. 4: t-SNE visualization of RoI features on (a) DUO and (b) RUOD validation sets. Left: baseline features are entangled and domain-stratified. Right: PhysFlowNet features are more class-compact and less domain-dependent. Colors denote object categories and marker shapes denote three K-Means clustered degradation domains.

Physical Interpretability of the Metric Tensor (G_{diag}). To intuitively interpret the learned metric, Fig. 5 shows representative examples from the TrashCan and Brackish datasets, including the ground-truth annotation, uncorrected activation, corrected activation after metric-guided Riemannian retraction, and metric deformation $\Delta G_{\mathrm{diag}} = \|G_{\mathrm{diag}} - \mathbf{I}\|$. Warmer colors indicate stronger activations or larger geometric adjustments. After retraction, responses become more concentrated around object regions, while background clutter is suppressed. Meanwhile, larger metric deformations often occur near objects or locally degraded areas, suggesting that G_{diag} adaptively applies stronger preconditioning where geometric rectification is needed. This demonstrates that metric-guided retraction improves object-centric representation and foreground–background separability.

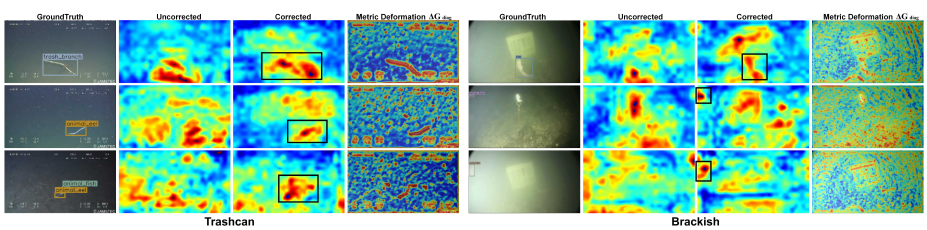

Fig. 5: Visualization of metric-guided feature rectification on the TrashCan (left) and Brackish (right) datasets. Columns show the ground truth, uncorrected activation, corrected activation after Riemannian retraction, and metric deformation $\Delta G_{\mathrm{diag}} = \|G_{\mathrm{diag}} - \mathbf{I}\|$. Warmer colors denote stronger activations or larger geometric adjustments. After retraction, responses become more object-centric with reduced background clutter.

Qualitative Evaluation of PMD. To assess the information-retention capability of PMD, we visualize activation maps at the final downsampling stage for an ablation setting. As shown in Fig. 6(c), the variant without PMD exhibits spatially spread activations, where the small target response is largely suppressed by surrounding clutter. In contrast, with PMD (Fig. 6(d)), the activation becomes noticeably more localized on the target region, while responses to scattering artifacts are attenuated, indicating that PMD improves boundary/region contrast and preserves target saliency during scale reduction.

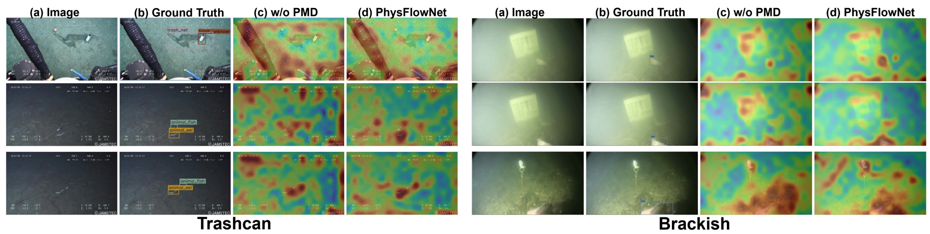

Fig. 6: Ablation of spatial feature activations on TrashCan and Brackish. (a) Raw images. (b) Ground truth. (c) Ours w/o PMD. (d) Ours with PMD. PMD effectively concentrates target activations and suppresses background clutter.

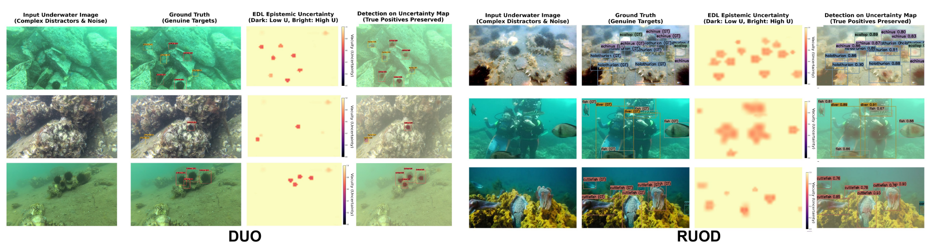

Fig. 7: Qualitative results of false-positive suppression via EDL. (a) Input. (b) Ground truth. (c) Multi-scale vacuity map, where brighter responses indicate higher epistemic uncertainty (weaker evidence) and darker responses indicate lower uncertainty (stronger evidence). (d) Detections overlaid on the vacuity map. PhysFlowNet localizes true targets in low-vacuity regions and suppresses detections in highly uncertain areas.

Qualitative Evaluation of Evidential Uncertainty. We visualize epistemic uncertainty (vacuity) to illustrate how the Dirichlet-based EDL module suppresses false positives induced by target-like underwater clutter (e.g., marine snow and light spots), where softmax-based detectors may be overconfident. Fig. 7(a–d) shows inputs, ground truth, the multi-scale vacuity map, and the corresponding detections on DUO and RUOD. True objects consistently align with low-vacuity (dark) regions, indicating sufficient evidence, whereas distractors yield high vacuity (bright), indicating weak or ambiguous evidence. Consequently, detections concentrate in low-vacuity areas, suggesting that EDL down-weights uncertain responses and reduces overconfident background detections. Combined with the Riemannian rectification that improves feature alignment, uncertainty-aware calibration further mitigates false positives in severely degraded scenes.

5 Conclusion

In this work, we present PhysFlowNet, a physics-conditioned manifold rectification framework that tackles underwater degradation from a Riemannian flow perspective. By modeling degradation as feature-space transport, we propose HPCE to leverage spatio-spectral physical priors to guide PPRB in performing stable metric-preconditioned retractions, effectively restoring distorted features to canonical semantics. Furthermore, we introduce PMD to mitigate cascaded smoothing during scale transitions via peak-preserving downsampling for robust object detection, and employ UECO for evidential uncertainty regulation to suppress backscatter-induced false positives in turbid scenes. Evaluations across four UOD benchmarks (DUO, Brackish, TrashCan, and RUOD) show that PhysFlowNet achieves state-of-the-art accuracy and efficiency, underscoring the immense potential of explicit geometric and physics-guided modeling for reliable detection in severely degraded environments.

Acknowledgements

This work was supported in part by the National Key Research and Development Program of China under Grant 2025YFE0218500; in part by the National Natural Science Foundation of China under Grant 62503139, Grant 62573163, and Grant 62521006; in part by Shenzhen Science and Technology Program under Grant JCYJ20240813105137049; and in part by the GuangDong Basic and Applied Basic Research Foundation under Grant 2024A1515012028, and Grant 2026A1515011822; and in part by the Shenzhen Medical Research Fund under Grant A2502034.

References

1. Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U.R., et al.: A review of uncertainty quantification in deep learning: Techniques, applications and challenges. Information fusion **76**, 243–297 (2021)
2. Akkaynak, D., Treibitz, T.: Sea-thru: A method for removing water from underwater images. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1682–1691 (2019)
3. Awad, A., Saleem, A., Paheding, S., Lucas, E., Al-Ratrout, S., Havens, T.C.: Beneath the surface: The role of underwater image enhancement in object detection. arXiv preprint arXiv:2411.14626 (2024)
4. Cai, Z., Vasconcelos, N.: Cascade r-cnn: Delving into high quality object detection. In: CVPR. pp. 6154–6162 (2018)

5. Chattopadhyay, P., Sarangmath*, K., Vijaykumar, V., Hoffman, J.: Pasta: Proportional amplitude spectrum training augmentation for syn-to-real domain generalization. In: IEEE/CVF International Conference in Computer Vision (ICCV) (2023)
6. Chen, G., Mao, Z., Shen, J., Cheng, Z.: Pseudo label guided object detection in sparsely annotated underwater optical images. IEEE Transactions on Geoscience and Remote Sensing (2025)
7. Chen, Y., Song, P., Liu, H., Dai, L., Zhang, X., Ding, R., Li, S.: Achieving domain generalization for underwater object detection by domain mixup and contrastive learning. Neurocomputing **528**, 20–34 (2023)
8. Dai, L., Liu, H., Song, P., Liu, M.: A gated cross-domain collaborative network for underwater object detection. PR **149**, 110222 (2024)
9. Dai, L., Liu, H., Song, P., Tang, H., Ding, R., Li, S.: Edge-guided representation learning for underwater object detection. CAAI Transactions on Intelligence Technology **9**(5), 1078–1091 (2024)
10. Davis, O., Kessler, S., Petrache, M., Ceylan, İ.İ., Bronstein, M., Bose, A.J.: Fisher flow matching for generative modeling over discrete data. Advances in Neural Information Processing Systems **37**, 139054–139084 (2024)
11. Finder, S.E., Amoyal, R., Treister, E., Freifeld, O.: Wavelet convolutions for large receptive fields. In: Proceedings of the European Conference on Computer Vision (2024)
12. Fu, C., Liu, R., Fan, X., Chen, P., Fu, H., Yuan, W., Zhu, M., Luo, Z.: Rethinking general underwater object detection: Datasets, challenges, and solutions. Neurocomputing **517**, 243–256 (2023)
13. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M., Lempitsky, V.: Domain-adversarial training of neural networks. Journal of machine learning research **17**(59), 1–35 (2016)
14. Gao, J., Chen, M., Xiang, L., Xu, C.: A comprehensive survey on evidential deep learning and its applications. IEEE Transactions on Pattern Analysis and Machine Intelligence **48**(3), 2118–2138 (2026)
15. Ge, H., Sun, P., Lu, Y.: A new dataset, model, and benchmark for lightweight and real-time underwater object detection. Neurocomputing **651**, 130891 (2025)
16. Goan, E., Fookes, C.: Bayesian Neural Networks: An Introduction and Survey, pp. 45–87. Springer International Publishing, Cham (2020)
17. Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q.: On calibration of modern neural networks. In: Precup, d., Teh, Y.W. (eds.) Proceedings of the 34th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 70, pp. 1321–1330. PMLR (06–11 Aug 2017)
18. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1956–1963 (2009)
19. Hong, J., Fulton, M., Sattar, J.: Trashcan: A semantically-segmented dataset towards visual detection of marine debris. CoRR **abs/2007.08097** (2020)
20. Islam, M.J., Xia, Y., Sattar, J.: Fast underwater image enhancement for improved visual perception. IEEE Robotics and Automation Letters **5**(2), 3227–3234 (2020)
21. Jocher, G., Chaurasia, A., Qiu, J.: Ultralytics yolov8 (2023)
22. Jocher, G., Qiu, J.: Ultralytics yolo11 (2024)
23. Kapoor, M., Patra, S., Subudhi, B.N., Jakhetiya, V., Bansal, A.: Underwater moving object detection using an end-to-end encoder-decoder architecture and graphsage with aggregator and refactoring. In: Proceedings of the IEEE/CVF Conference

on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 5636–5645 (June 2023)
24. Khan, A., Fouda, M.M., Do, D.T., Almaleh, A., Alqahtani, A.M., Rahman, A.U.: Underwater target detection using deep learning: Methodologies, challenges, applications, and future evolution. IEEE Access **12**, 12618–12635 (2024)
25. Li, C., Anwar, S., Hou, J., Cong, R., Guo, C., Tao, D.: Underwater image enhancement via medium transmission-guided multi-color space embedding. IEEE Transactions on Image Processing **30**, 4985–5000 (2021)
26. 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 **29**, 4376–4389 (2020)
27. Li, T., He, K.: Back to basics: Let denoising generative models denoise. arXiv preprint arXiv:2511.13720 (2026)
28. Liang, X., Song, P.: Excavating roi attention for underwater object detection. In: Proceedings of the IEEE International Conference on Image Processing. IEEE (2022)
29. Lin, H., Liang, D., Qi, Z., Bai, X.: A unified image-dense annotation generation model for underwater scenes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2025)
30. Lin, P., Wang, Y., Li, Y., Yang, Y.R., Fu, X.: A real-world underwater turbid image enhancement benchmark and beyond. Knowledge-Based Systems **330**, 114683 (2025)
31. Lin, T., Zha, H.: Riemannian manifold learning. IEEE Transactions on Pattern Analysis and Machine Intelligence **30**(5), 796–809 (2008)
32. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 2999–3007 (2017)
33. Liu, C., Li, H., Wang, S., Zhu, M., Wang, D., Fan, X., Wang, Z.: A dataset and benchmark of underwater object detection for robot picking. In: 2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). pp. 1–6 (2021)
34. Liu, R., Jiang, Z., Yang, S., Fan, X.: Twin adversarial contrastive learning for underwater image enhancement and beyond. TIP **31**, 4922–4936 (2022)
35. Liu, W., Pang, J., Zhang, B., Wang, J., Liu, B., Tao, D.: See degraded objects: A physics-guided approach for object detection in adverse environments. IEEE Transactions on Image Processing (2025)
36. Liu, X., Li, C., Guo, S.: Wfda: Wavelet-based frequency decomposition and aggregation for underwater object detection. In: 2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 15127–15134 (2025)
37. Long, H., Shen, L., Wang, Z., Chen, J.: Underwater forward-looking sonar images target detection via speckle reduction and scene prior. IEEE Transactions on Geoscience and Remote Sensing **61**, 1–13 (2023)
38. Long, M., Cao, Z., Wang, J., Jordan, M.I.: Conditional adversarial domain adaptation. Advances in neural information processing systems **31** (2018)
39. Lucas, E., Awad, A., Geglio, A., Moradi, S., Saleem, A., Havens, T., Galloway, A., Paheding, S.: Underwater image enhancement and object detection: Are poor object detection results on enhanced images due to missing human labels? In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW). pp. 1520–1525 (2025)
40. McGlamery, B.L.: A Computer Model For Underwater Camera Systems. In: Duntley, S.Q. (ed.) Ocean Optics VI. vol. 0208, pp. 221 – 231. International Society for Optics and Photonics, SPIE (1980)

41. Munir, M.A., Khan, M.H., Khan, S., Khan, F.S.: Bridging precision and confidence: A train-time loss for calibrating object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11474–11483 (June 2023)
42. Park, D., Lee, T., Joo, M., Kim, H.J.: Blockwise flow matching: Improving flow matching models for efficient high-quality generation. arXiv preprint arXiv:2510.21167 (2025)
43. Pedersen, M., Haurum, J.B., Gade, R., Moeslund, T.B., Madsen, N.: Detection of marine animals in a new underwater dataset with varying visibility. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 18–26 (June 2019)
44. Qiao, S., Chen, L.C., Yuille, A.: DetectoRS: Detecting objects with recursive feature pyramid and switchable atrous convolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10213–10224 (June 2021)
45. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. TPAMI **39**(6), 1137–1149 (2017)
46. Song, P., Li, P., Dai, L., Wang, T., Chen, Z.: Boosting r-cnn: Reweighting r-cnn samples by rpn's error for underwater object detection. Neurocomputing **530**, 150–164 (2023)
47. Walker, J.L., Zeng, Z., Wu, C.L., Jaffe, J.S., Frasier, K.E., Sandin, S.S.: Underwater object detection under domain shift. IEEE Journal of Oceanic Engineering (2024)
48. 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. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 7464–7475 (2023)
49. Wang, W., Li, H., Ding, Z., Nie, F., Chen, J., Dong, X., Wang, Z.: Rethinking maximum mean discrepancy for visual domain adaptation. IEEE Transactions on Neural Networks and Learning Systems **34**(1), 264–277 (2023)
50. Wang, Z., Shen, L., Xu, M., Yu, M., Wang, K., Lin, Y.: Domain adaptation for underwater image enhancement. IEEE Transactions on Image Processing **32**, 1442–1457 (2023)
51. Xu, G., Liao, W., Zhang, X., Li, C., He, X., Wu, X.: Haar wavelet downsampling: A simple but effective downsampling module for semantic segmentation. Pattern Recognition **143**, 109819 (2023)
52. Xu, H., He, J., Pang, C., Yu, Y.: Sfudnet: Underwater object detection via spatial-frequency domain modulation with mixture of experts. Knowledge-Based Systems **324**, 113805 (2025)
53. Xu, J., Li, W., Sun, H., Li, F., Wang, Z., Peng, L., Ren, J., Yang, H., Hu, X., Pei, R., et al.: Fast image super-resolution via consistency rectified flow. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 11755–11765 (2025)
54. Xu, Q., Zhang, R., Zhang, Y., Wang, Y., Tian, Q.: A fourier-based framework for domain generalization. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 14378–14387 (2021)
55. Zhang, D., Yu, C., Li, Z., Qin, C., Xia, R.: A lightweight network enhanced by attention-guided cross-scale interaction for underwater object detection. Applied Soft Computing **184**, 113811 (2025)
56. Zhang, H., Li, S., Wang, P., Zeng, D., Ge, S.: M3d: Dataset condensation by minimizing maximum mean discrepancy. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 9314–9322 (2024)

57. Zhang, W., Wang, H., Li, H., Li, Y., Song, T., Liu, H., Wang, G.: Dual-stream feature pyramid network with task interaction for underwater object detection. Digital Signal Processing **163**, 105199 (2025)
58. Zheng, Y., Chen, W., Lin, R., Zhao, T., Le Callet, P.: Uif: An objective quality assessment for underwater image enhancement. IEEE Transactions on Image Processing **31**, 5456–5468 (2022)
59. Zhou, J., He, Z., Lam, K.M., Wang, Y., Zhang, W., Guo, C., Li, C.: Amsp-uod: When vortex convolution and stochastic perturbation meet underwater object detection. In: AAAI. vol. 38, pp. 7659–7667 (2024)
60. Zhou, J., He, Z., Zhang, D., Liu, S., Fu, X., Li, X.: Spatial residual for underwater object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence **47**(6), 4996–5013 (2025)
61. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV) (2017)