Journal of Computer Science and Cybernetics, V.42, N.2 (2026), 103–118
DOI no: 10.15625/1813-9663/23647

ENERGY DISTANCE-BASED PANOPTIC SEGMENTATION FOR UNSUPERVISED ANOMALY DETECTION IN AQUATIC ENVIRONMENTS

TOAN PHUNG HUYNH[1,2], TAI VAN VO[1], HIEP XUAN HUYNH[1,2,3,*]

[1]*Can Tho University (CTU), Campus II, 3/2 Street, Ninh Kieu Ward, Can Tho City, Viet Nam*
[2]*CTU Leading Research Team on Automation, Artificial Intelligence, Information Technology and Digital Transformation (CTU-AIMED), Campus II, 3/2 Street, Ninh Kieu Ward, Can Tho City, Viet Nam*
[3]*CTU Key Research Team on Sustainable Aquaculture Development and Climate Change Adaptation, Campus II, 3/2 Street, Ninh Kieu Ward, Can Tho City, Viet Nam*

Abstract. Dead fish detection in industrial pangasius catfish ponds is critical for preventing disease spread and economic loss, yet manual monitoring is labor-intensive and impractical at scale. Existing computer vision approaches rely on supervised deep learning, which requires large labeled datasets that are prohibitively difficult to collect for rare mortality events in complex aquatic environments. We propose an unsupervised framework that casts dead fish detection as an Energy Distance-based Panoptic Segmentation problem. The system computes a 12-dimensional feature map encoding intensity, edge, color, contrast, and water-mask information, then derives an Energy Distance Map (ED Map) via a weighted sliding-window strategy. A two-step panoptic procedure separates low-energy regions as live-fish instances (*things*) from high-energy regions as anomalous dead-fish instances (*things*), while water pixels remain as the *stuff* class. Final classification employs a dynamic 75^{th} percentile threshold, requiring no training data. Experiments on three real industrial pangasius pond scenarios - high-density with non-uniform lighting, medium-density with stable lighting, and low-density with high contrast—yield ROC-AUC values of 0.953, 0.914, and 0.855, and Average Precision of 0.420, 0.334, and 0.114, demonstrating strong anomaly localization across diverse operational conditions.

Keywords. Panoptic segmentation, energy distance, energy-based model, point of interest, dead catfish detection, unsupervised anomaly detection, aquaculture.

1. INTRODUCTION

The aquaculture industry plays a crucial role in food security and economic development, especially in Asian countries [1]. Vietnam is one of the world's largest pangasius catfish exporters, with production exceeding 1.5 million tons annually and export value surpassing 2 billion USD [2]. In the context of developing high-tech agriculture (Agriculture 4.0),

*Corresponding author.
E-mail addresses: hptoan@ctu.edu.vn (T.P. Huynh); vvtai@ctu.edu.vn (T.V. Vo); hxhiep@ctu.edu.vn (H.X. Huynh).

applying artificial intelligence and computer vision to pond management has become an inevitable trend [3]. Particularly in intensive pangasius farming models with high density (150–200 fish/m^3), monitoring fish health and early detection of abnormal signs is extremely important [4]. Dead fish phenomena in ponds can occur due to various causes such as oxygen deficiency, disease, or poisoning, and if not detected promptly will lead to disease spread, water pollution, and serious economic losses [5, 6]. Traditional monitoring methods based on manual observation have many limitations in efficiency, continuity, and accuracy, while consuming considerable labor [7]. Therefore, developing automatic dead fish detection systems is an urgent need in production practice.

Automatic dead fish detection in real pond environments poses significant technical challenges. Pond environments have complex characteristics with non-uniform lighting conditions due to water surface reflection, shadows, and natural light changes [8]. Water turbidity, high color similarity between fish and environment, dense fish populations causing mutual occlusion, along with diversity in fish postures create a challenging computer vision problem [9, 10]. In recent years, deep learning methods such as YOLO, Faster R-CNN, and Mask R-CNN have been successfully applied in fish detection and segmentation with high accuracy [11, 12, 13]. However, these supervised learning methods all require large training datasets with thousands of accurately labeled images and need powerful computational resources [14]. Collecting and labeling dead fish data in real conditions is very difficult due to the rare and uncontrollable nature of this event [15]. Meanwhile, traditional image processing techniques such as thresholding, edge detection, and clustering, although not requiring training data, are often unstable in complex environments [16, 17]. Therefore, an unsupervised method capable of adapting to real conditions without requiring large labeled datasets is an urgent need.

Panoptic segmentation, proposed by Kirillov et al. [18], unifies Semantic Segmentation (*stuff* classes: amorphous background regions) and Instance Segmentation (*things* classes: countable object instances), assigning every pixel both a semantic class label and an instance identifier. This unified representation makes panoptic segmentation a natural fit for aquatic scenes where the water background is *stuff* and individual fish are *things*. However, existing panoptic methods are deep learning-based and require large labeled datasets [19, 20, 21]. Energy distance [22, 23] provides a label-free measure of feature-distribution divergence between image regions, offering a principled way to identify anomalous regions without supervision. Combining panoptic segmentation with energy distance in an unsupervised framework for dead fish detection is precisely the research gap this study addresses.

This study proposes an automatic dead pangasius catfish detection system based on unsupervised learning, combining panoptic segmentation with energy distance and Point of Interest (PoI).

Three fundamental distinctions separate this work from the existing literature. First, dead fish detection is formulated as an Energy Distance-based Panoptic Segmentation problem — an original coupling of statistical divergence theory with instance-level scene understanding that has not been explored in prior aquaculture vision research. Second, the proposed framework operates without any labeled training data: anomalous regions are identified solely through feature-distribution divergence computed on the test image itself, making the system deployable in real farms where mortality-event annotation is prohibitively costly.

Third, the entire pipeline runs on standard CPU hardware with linear $\mathcal{O}(N)$ computational complexity, eliminating the GPU infrastructure required by deep learning baselines and enabling continuous real-time monitoring on resource-constrained farm deployments.

Main contributions include: (1) A theoretical model grounding the panoptic label map and Energy Distance computation within an Energy-Based Model (EBM) framework; (2) A 12-dimensional feature extraction specialized for aquatic environments, covering intensity, edge, color, contrast, and water-environment mask; (3) An Energy Distance Map algorithm with a weighted sliding-window strategy, where high-energy regions are hypothesized as potential dead fish; (4) A two-step panoptic segmentation: low-energy instances (live fish) as *things* and high-energy instances (suspected dead fish) as anomalous *things*, with water as *stuff*; (5) Experimental validation on three real industrial pangasius pond scenarios, achieving ROC-AUC ≥ 0.855 across all conditions.

The remainder of the paper is organized as follows: Section 2 presents related works; Section 3 develops the theoretical model; Section 4 describes the functional model; Section 5 reports experiments and results; Section 6 draws conclusions and outlines future directions.

2. RELATED WORKS

Computer vision in aquaculture monitoring. Applying computer vision to smart fish farming has been comprehensively reviewed [7], identifying challenges of data scarcity, complex lighting, and real-time requirements. Non-uniform illumination, turbidity, and surface reflection in pond imagery are extensively studied [9], and early disease detection via visual monitoring is critical to prevent epidemic spread [6]. Recent work has also investigated deep learning approaches for real-time fish behavior recognition in pond monitoring [24].

Supervised deep learning for fish detection. CNN-based methods achieve state-of-the-art accuracy in fish detection: hybrid motion-learning on underwater videos [11], preprocessing-augmented recognition [13], and deep counting surveys [12]. However, supervised methods require thousands of labeled images [14] — a critical barrier for dead fish detection where mortality annotations are extremely scarce [15].

Unsupervised, label-free, and panoptic approaches. Label-free methods include local-descriptor feature learning [16], zero-shot CLIP-based anomaly localization [25], diffusion-model anomaly detection [26], and neuro-fuzzy behavior monitoring [17]. Panoptic segmentation [18] provides unified *stuff*/*things* representation; strong supervised baselines include Panoptic-DeepLab [20] and EfficientPS [21]. Energy distance [22, 23] offers non-parametric, label-free divergence scoring. No prior work combines panoptic segmentation with energy distance in an unsupervised dead fish detection framework — the central contribution of this study.

3. THEORETICAL MODELING

3.1. Panoptic segmentation formalism

Following [18], a panoptic label map over image domain $\Omega = \{1, \ldots, H\} \times \{1, \ldots, W\}$ assigns each pixel (x, y) a pair (c, id), where $c \in \mathcal{C}$ is a semantic class and id is an instance identifier. The semantic class set $\mathcal{C}$ is partitioned into **stuff** $\mathcal{C}_s$ and **things** $\mathcal{C}_t$. Stuff classes represent amorphous background regions with no countable instances (e.g., water, sky); all

pixels in a stuff class share a single semantic label with $id = 0$. Things classes represent countable object instances with clear boundaries (e.g., individual fish); each detected instance k receives a unique identifier $id = k \geq 1$.

Definition 1. A panoptic label map is a function $\mathcal{L} : \Omega \to \mathbb{Z}_{\geq 0}$ where

$$\mathcal{L}(x, y) = \begin{cases} 0 & \text{if pixel } (x, y) \text{ belongs to the stuff class (water/background),} \\ k \geq 1 & \text{if pixel } (x, y) \text{ belongs to things instance } k. \end{cases} \tag{1}$$

In our aquatic context, we define two classes: $\mathcal{C}_s = \{\text{water}\}$ (stuff) and $\mathcal{C}_t = \{\text{live fish, dead fish}\}$ (things). The system produces $\mathcal{L}$ without supervision, using Energy Distance to separate stuff from things and to label anomalous instances as dead fish.

3.2. Energy-based model framework

Energy-based models (EBMs), introduced by LeCun et al. [27], provide a unifying framework for supervised, unsupervised, and self-supervised learning. An EBM defines a scalar energy function $\mathcal{F} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ that measures the compatibility between an input x and an output y. Inference finds the output $\check{y}$ that minimizes the energy

$$\check{y} = \underset{y}{\operatorname{argmin}}\ \mathcal{F}(x, y). \tag{2}$$

via the Gibbs-Boltzmann distribution, $P(y \mid x) \propto \exp(-\beta \mathcal{F}(x, y))$ where $\beta > 0$ is an inverse temperature [27]. Crucially, EBMs do not require explicit normalization, making them tractable for high-dimensional output spaces.

In our framework, the Energy Distance Map $\mathrm{ED}(x, y)$ serves as the pixelwise energy function $\mathcal{F}$. A pixel (x, y) is assigned high energy if its local feature distribution diverges strongly from neighboring regions. Anomaly detection reduces to

$$\hat{y}(x, y) = \begin{cases} \text{anomalous (dead fish)} & \text{if } \mathrm{ED}(x, y) \geq \tau_{\text{high}} \\ \text{normal} & \text{otherwise.} \end{cases} \tag{3}$$

The computation graph of this EBM-inspired system is illustrated in Figure 1.

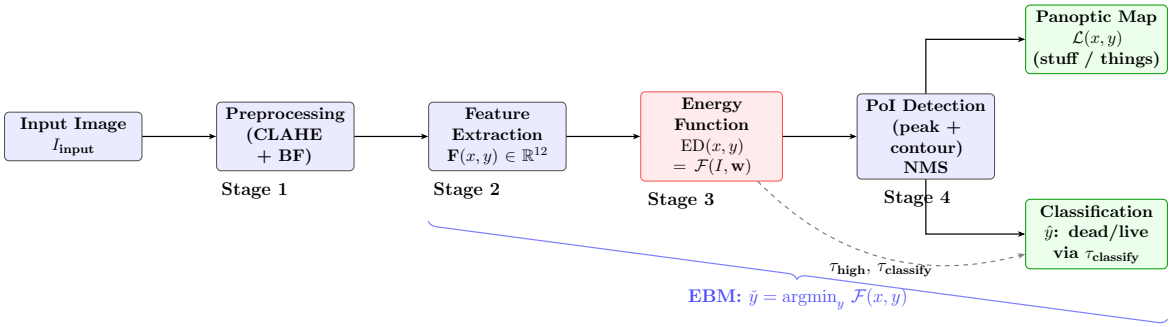

Figure 1: EBM computation graph. The Energy Distance Map $\mathrm{ED}(x, y)$ plays the role of the energy function $\mathcal{F}$; the panoptic map $\mathcal{L}$ assigns stuff (water, $\mathcal{L}$=0) and things (fish instances, $\mathcal{L}$=k).

3.3. Energy distance and feature space

Let $X \sim \mathbb{P}$ and $Y \sim \mathbb{Q}$ be random vectors in $\mathbb{R}^d$. The energy distance [22, 23] is

$$\mathcal{E}^2(\mathbb{P}, \mathbb{Q}) = 2\mathbb{E}\|X - Y\| - \mathbb{E}\|X - X'\| - \mathbb{E}\|Y - Y'\| \geq 0 \tag{4}$$

with equality iff $\mathbb{P} = \mathbb{Q}$ (Theorem 1 [22]), making $\mathcal{E}^2$ a faithful divergence: pixels whose feature distribution diverges from neighbors receive strictly positive energy.

The feature map $\mathbf{F} : \Omega \to \mathbb{R}^{12}$ covers five modalities: intensity (f_1), edges (f_2, f_3, f_4), local contrast (f_5), water mask (f_6), HSV color (f_7–f_9), RGB color (f_{10}–f_{12}), each normalized to $[0, 1]$. Weight vector $\mathbf{w} = [1.5, 2.0, 1.0, 1.5, 2.0, 0.5, 1.0, 1.0, 1.0, 1.2, 1.2, 1.2]^T$ up-weights edge/contrast (w_2=w_5=2.0) and down-weights water mask (w_6=0.5). For window size κ and h=$\lfloor \kappa/2 \rfloor$, the pixelwise ED is computed via center/neighbor means $\boldsymbol{\mu}_c$, $\boldsymbol{\mu}_n$

$$\delta(x, y, d) = \|\mathbf{w} \odot (\boldsymbol{\mu}_c(x, y) - \boldsymbol{\mu}_n(x, y, d))\|_2, \tag{5}$$

$$\mathrm{ED}(x, y) = \frac{1}{|D|} \sum_{d \in D} \delta(x, y, d), \tag{6}$$

instantiating Eq. (4) in the finite-window setting where δ approximates $\mathbb{E}\|X-Y\|$ via weighted Euclidean distance of empirical means.

4. FUNCTIONAL MODELING

4.1. System overview

The proposed system executes five sequential stages: (1) pond image preprocessing, (2) 12-dimensional feature extraction, (3) Energy Distance Map computation, (4) PoI-based candidate detection, and (5) two-step panoptic segmentation and classification. The overall architecture is illustrated in Figure 2.

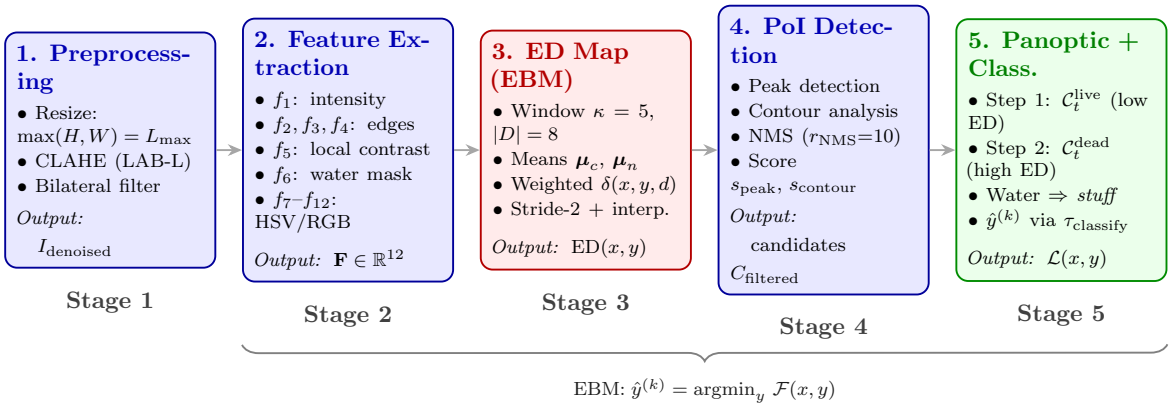

Figure 2: System architecture: five-stage pipeline from raw pond image to panoptic dead-fish map. Blue = image processing stages; red = energy function (EBM core); green = panoptic segmentation and classification output.

The core hypothesis is that dead fish exhibit feature distributions significantly diverging from live fish and water, manifested through lighter color (oxygen deficiency), different surface

texture (floating/reflecting), and high local contrast. Consequently, dead fish pixels receive high ED values, enabling unsupervised separation from the background.

4.2. Preprocessing function

The preprocessing stage standardizes input images and enhances contrast to facilitate feature extraction. The input image I_{input} is first resized so that $\max(H, W) = L_{\max} = 800\,\text{px}$, reducing computational load while preserving spatial detail. Contrast is then enhanced by applying CLAHE to the L channel in LAB color space (clipLimit=3.0, tileSize=8×8), which equalizes local contrast non-uniformity caused by water-surface reflections. Finally, a bilateral filter (9×9, $\sigma_s = \sigma_r = 75$) suppresses photon noise while preserving fish boundaries, yielding the denoised image I_{denoised}.

4.3. Feature extraction function

Feature extraction instantiates the theoretical model of Section 3.4, computing all 12 features from I_{denoised}. Throughout, $\varepsilon = 10^{-6}$ prevents division by zero.

The **intensity feature** f_1 is the luminance channel normalized to $[0, 1]$

$$f_1(x, y) = \frac{0.299\,R + 0.587\,G + 0.114\,B}{255}. \tag{7}$$

Edge features f_2 and f_3 capture Sobel gradient magnitude and orientation respectively, providing structural boundary cues that are highly discriminative for fish contours

$$f_2 = \frac{\sqrt{G_x^2 + G_y^2}}{\max_{(x,y)} \sqrt{G_x^2 + G_y^2} + \varepsilon}, \quad f_3 = \frac{\arctan 2(G_y, G_x)}{\pi}. \tag{8}$$

The **Laplacian feature** f_4 detects rapid intensity changes associated with dead fish surface texture

$$f_4 = \frac{\nabla^2 f_1}{\max_{(x,y)} |\nabla^2 f_1| + \varepsilon}. \tag{9}$$

Local contrast f_5 is computed via a Laplacian-of-Gaussian-like 3×3 kernel K_c, emphasizing center-surround differences characteristic of floating fish

$$f_5 = \frac{|K_c * f_1|}{\max_{(x,y)} |K_c * f_1| + \varepsilon}, \quad K_c = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}. \tag{10}$$

The **water mask** f_6 serves as the *stuff* indicator in the panoptic formalism. It is a binary flag obtained by HSV thresholding

$$f_6(x, y) = \mathbf{1}[\tilde{h}_l \le \tilde{h} \le \tilde{h}_u,\ \tilde{s}_l \le \tilde{s} \le \tilde{s}_u,\ \tilde{v}_l \le \tilde{v} \le \tilde{v}_u] \tag{11}$$

with $(\tilde{h}_l, \tilde{h}_u) = (40, 120)$, $(\tilde{s}_l, \tilde{s}_u) = (30, 255)$, $(\tilde{v}_l, \tilde{v}_u) = (30, 200)$, so that pixels with $f_6 = 1$ are *stuff* candidates. Finally, the six **color features** f_7–f_{12} encode the full HSV and RGB color spaces, each normalized to $[0, 1]$

$$f_7 = \tilde{h}/179, \quad f_8 = \tilde{s}/255, \quad f_9 = \tilde{v}/255, \tag{12}$$

$$f_{10} = R/255, \quad f_{11} = G/255, \quad f_{12} = B/255. \tag{13}$$

Algorithm 1 Preprocessing and 12-dimensional feature extraction

Require: I_{input}, $L_{\max}$=800, ε=10^{-6}
Ensure: Feature map $\mathbf{F} \in \mathbb{R}^{H\times W\times 12}$
$I_r \leftarrow \text{Resize}(I_{\text{input}},\ s{=}L_{\max}/\max(H,W))$; CLAHE on LAB-L; BilateralFilter $\rightarrow I_{\text{denoised}}$
$f_1 \leftarrow (0.299R{+}0.587G{+}0.114B)/255$; Sobel $\rightarrow G_x, G_y$
$f_2 \leftarrow \|G\|/(\max\|G\|{+}\varepsilon)$; $f_3 \leftarrow \arctan 2(G_y,G_x)/\pi$; $f_4 \leftarrow \nabla^2 f_1/(\max|\nabla^2 f_1|{+}\varepsilon)$
$f_5 \leftarrow |K_c{*}f_1|/(\max|K_c{*}f_1|{+}\varepsilon)$; $f_6 \leftarrow \mathbf{1}[\tilde{h}_l{\le}\tilde{h}{\le}\tilde{h}_u,\ \tilde{s}_l{\le}\tilde{s}{\le}\tilde{s}_u,\ \tilde{v}_l{\le}\tilde{v}{\le}\tilde{v}_u]$
$f_7{=}\tilde{h}/179$, $f_8{=}\tilde{s}/255$, $f_9{=}\tilde{v}/255$, $f_{10}{=}R/255$, $f_{11}{=}G/255$, $f_{12}{=}B/255$
$\mathbf{F}(x,y) \leftarrow [f_1,\ldots,f_{12}]^\top$
return F

4.4. Energy distance map function

The ED Map function instantiates Eqs. (5)–(6) with $\kappa = 5$ (hence $h = 2$). To reduce computational cost, ED is computed at stride of 2, and intermediate values are linearly interpolated. The output ED : $\Omega \to [0,1]$ serves as the pixelwise anomaly score: high ED indicates strong divergence from the local neighborhood, consistent with the EBM interpretation in Section 3.2.

Algorithm 2 Energy distance map computation

Require: $\mathbf{F}$, $\mathbf{w}$, κ=5, stride s_t=2, $|D|$=8
Ensure: $\text{ED} \in [0,1]^{H\times W}$
$h \leftarrow \lfloor \kappa/2 \rfloor$; init $\text{ED}_s \leftarrow \mathbf{0}$
for each (x,y) at stride s_t **do**
  $\boldsymbol{\mu}_c \leftarrow \frac{1}{\kappa^2}\sum_{i,j=-h}^{h} \mathbf{F}(x{+}i, y{+}j)$
  $\text{ED}_s(x,y) \leftarrow \frac{1}{|D|}\sum_{d\in D} \|\mathbf{w}\odot(\boldsymbol{\mu}_c - \boldsymbol{\mu}_n^d)\|_2$
end for
$\text{ED} \leftarrow \text{BilinearInterp}(\text{ED}_s)/(\max \text{ED}_s{+}\varepsilon)$
return ED

4.5. PoI detection function

Two complementary strategies detect Point of Interest (PoI) candidates corresponding to fish locations.

Strategy 1, peak detection. The ED map is smoothed with a 3×3 Gaussian filter. Local maxima above the 70th-percentile threshold τ_{dynamic} are detected with minimum inter-peak distance $r = \max(5, \lfloor r_{\text{NMS}}/2 \rfloor)$. Each peak (x,y) receives a score

$$s_{\text{peak}}(x,y) = \text{ED}(x,y)\cdot(1 + f_2(x,y) + f_5(x,y)). \tag{14}$$

Strategy 2, contour analysis. The combined edge-contrast feature $f_{\text{combined}} = (f_2 + f_5)/2$ is thresholded at its 75th percentile. After morphological cleaning (5×5 ellipse kernel), contours with area $\in [A_{\min}, A_{\max}]$ are extracted. Each contour centroid (x_c, y_c) receives a score

$$s_{\text{contour}}(x_c,y_c) = \text{ED}(x_c,y_c)\cdot s_{\text{shape}}\cdot 2.0, \tag{15}$$

where $s_{\text{shape}} \in \{0.5, 1.0\}$ based on circularity.

Both candidate sets are merged and filtered by Non-Maximum Suppression (see Algorithm 4).

Algorithm 3 PoI detection, panoptic segmentation, and classification

Require: ED, $\mathbf{F}$, r_{NMS}, $[A_{\min}, A_{\max}]$, σ_{PoI}=25, θ=0.3
Ensure: Labels $\{\hat{y}^{(k)}\}$, panoptic map $\mathcal{L}$
// PoI Detection
$C_{\text{peak}} \leftarrow$ local maxima of GaussFilter(ED) above 70th-pct, scored by s_{peak}=ED$\cdot(1+f_2+f_5)$
$f_c \leftarrow (f_2+f_5)/2$; extract contours of $(f_c>\tau_{75})$ with area $\in[A_{\min}, A_{\max}]$, scored by s_c=ED$\cdot s_{\text{shape}}\cdot 2$
$C_{\text{filtered}} \leftarrow \text{NMS}(C_{\text{peak}} \cup C_{\text{contour}},\ r_{\text{NMS}})$ ▷ Alg. 4
// Step 1: Live fish (things $\mathcal{C}_t^{\text{live}}$, low energy)
$\mathcal{L} \leftarrow \mathbf{0}$, $k \leftarrow 0$; build W_{PoI} from top-20 candidates (Gaussian σ_{PoI})
$M_{\text{bin}} \leftarrow (f_c>\tau_{75})$; morphological clean; label components R with area $\in[A_{\min}, A_{\max}]$: $\mathcal{L}(R)\leftarrow k{+}{+}$
// Step 2: Dead fish (things $\mathcal{C}_t^{\text{dead}}$, high energy)
$M_{\text{high}} \leftarrow (\text{ED}\geq\tau_{\text{high}})$, τ_{high}=Pct(ED, 90); dilate 7×7 ellipse
for each component R' with area $\in[A_{\min}, A_{\max}]$ **do**
  if $|\{R' : \mathcal{L}>0\}|/|R'| < \theta$ **then** $\mathcal{L}(R')\leftarrow k{+}{+}$
  end if
end for
// Classification
$\tau_c \leftarrow \text{Pct}(\{\text{ED} : \mathcal{L}>0\}, 75)$; $\hat{y}^{(k)} \leftarrow$ "Dead" if $E_{\text{avg}}^{(k)}\geq\tau_c$ else "Live"
return $\mathcal{L}$, $\{\hat{y}^{(k)}\}$

4.6. Two-step panoptic segmentation function

The panoptic label map $\mathcal{L}$ (Definition 1) is constructed in two steps that mirror the panoptic formalism: Step 1 creates *things* instances for normal fish; Step 2 creates *things* instances for anomalous (dead) fish. Pixels not assigned to any instance remain as *stuff* (water, $\mathcal{L} = 0$).

Step 1, normal fish instances (low energy, things $\mathcal{C}_t^{\text{live}}$). A PoI weight map W_{PoI} is constructed from the top-20 scoring candidates using Gaussian weights ($\sigma_{\text{PoI}} = 25$). A binary mask from f_{combined}

$$M_{\text{binary}}(x, y) = \mathbf{1}[f_{\text{combined}}(x, y) > \tau_{75}], \quad \tau_{75} = \text{Percentile}(f_{\text{combined}}, 75), \tag{16}$$

is morphologically cleaned. Connected components with $A_{\min} \leq \text{area} \leq A_{\max}$ receive unique labels $k = 1, 2, \ldots$

Step 2, dead fish instances (high energy, $\mathcal{C}_t^{\text{dead}}$). High-energy mask using the 90th-percentile threshold

$$M_{\text{high}}(x, y) = \mathbf{1}[\text{ED}(x, y) \geq \tau_{\text{high}}], \quad \tau_{\text{high}} = \text{Percentile}(\text{ED}, 90), \tag{17}$$

is cleaned with a 7×7 ellipse kernel. Each high-energy component with an overlap ratio < 0.3 receives a new label k; its mean energy $E_{\text{avg}}^{(k)}$ is recorded. The final panoptic map $\mathcal{L}(x, y) \in \{0, 1, \ldots, N_{\text{inst}}\}$ satisfies Definition 1: $\mathcal{L} = 0$ for water pixels (*stuff*); $\mathcal{L} = k \geq 1$ for fish instances (*things*).

4.7. Classification function

Each instance k is classified based on its average energy relative to a global adaptive threshold

$$\tau_{\text{classify}} = \text{Percentile}(\{\text{ED}(x,y) : \mathcal{L}(x,y) > 0\},\ 75)\,, \tag{18}$$

$$E_{\text{avg}}^{(k)} = \frac{\sum_{(x,y):\,\mathcal{L}(x,y)=k} \text{ED}(x,y)}{\#\{(x,y) :\ \mathcal{L}(x,y) = k\}}, \tag{19}$$

$$\hat{y}^{(k)} = \begin{cases} \text{“Dead fish”} & \text{if } E_{\text{avg}}^{(k)} \geq \tau_{\text{classify}}, \\ \text{“Live fish”} & \text{otherwise.} \end{cases} \tag{20}$$

This adaptive threshold requires no labeled data and adjusts automatically to the energy distribution of each image. Algorithm 3 provides the complete pseudo-code for PoI detection, panoptic segmentation, and classification.

4.8. Theoretical remark: strengths and limitations

Strengths: (1) No labeled data - Energy Distance is computed directly from pixel features; (2) Linear $\mathcal{O}(N)$ complexity enabling CPU real-time processing (≈2–4 s/image); (3) Principled EBM grounding [27]: anomaly detection reduces to energy minimization; (4) Percentile-based adaptive thresholds (τ_{high}, τ_{classify}) self-calibrate per image without tuning; (5) Panoptic *stuff*/*things* separation yields structured, interpretable output.
Limitations: (1) Performance degrades under extreme turbidity where fish/water contrast disappears; (2) Hand-crafted 12-D features are less expressive than deep CNN features for fine-grained variation; (3) Post-hoc ground-truth mask generation from high-energy regions may introduce evaluation bias; (4) ED is sensitive to window size κ: small κ misses large dead fish, large κ blurs boundaries.

5. EXPERIMENTS

5.1. Data used

Data were collected from real industrial pangasius catfish ponds in the Mekong Delta region, Viet Nam. Three representative images spanning diverse density and lighting conditions were selected. Images were captured using high-resolution cameras under natural lighting; raw resolutions exceed 2000×1500 pixels, normalized to 800×452 during preprocessing. The method is fully unsupervised with no training/validation/test split; ground-truth dead-fish pixel masks were generated post-hoc from detected high-energy panoptic instances for pixel-level ROC-AUC and Average Precision (AP) evaluation.

5.2. Tool used

The system is implemented in Python 3.10 using OpenCV 4.8 (image processing), NumPy 1.24 (array computation), and SciPy 1.11 (peak detection). No deep learning frameworks or GPU are required; the pipeline runs on standard CPU hardware (Intel Core i7, 16 GB RAM, Ubuntu 22.04 LTS) at approximately 2–4 seconds per image. All parameters (Table 1) are fixed uniformly across all three scenarios without per-image tuning.

Table 1: System parameters used in all experiments

Parameter	Value	Description
$L_{\max}$	800 px	Max image dimension
κ	5	ED sliding window size
r_{NMS}	10 px	NMS suppression radius
σ_{PoI}	25.0	Gaussian weight sigma
$A_{\min}$	200 px^2	Min instance area
$A_{\max}$	10 000 px^2	Max instance area
p_{high}	90th pct	High-energy threshold percentile
p_{classify}	75th pct	Classification threshold percentile

Algorithm 4 formalizes the Non-Maximum Suppression step: candidates are sorted by score and suppressed if within $r_{\text{NMS}} = 10$ px of a higher-scoring selection.

Algorithm 4 Non-Maximum Suppression (NMS)

Input: Candidates C (each as (x, y, s, m): position, score, mode$\in$\{peak, contour\}), radius r_{NMS}
Output: Filtered candidates C_{filtered}

$C_{\text{filtered}} \leftarrow \emptyset$
$C_{\text{sorted}} \leftarrow \text{sort}(C \text{ by } s \text{ descending})$
for each $(x_i, y_i, s_i, m_i) \in C_{\text{sorted}}$ **do**
    dup $\leftarrow$ **false**
    for each $(x_j, y_j, s_j, m_j) \in C_{\text{filtered}}$ **do**
        if $\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} < r_{\text{NMS}}$ **then**
            dup $\leftarrow$ **true**
            if $s_i > s_j$ **then**
                replace (x_j, y_j, s_j, m_j) with (x_i, y_i, s_i, m_i)
            end if
            break
        end if
    end for
    if not dup **then** $C_{\text{filtered}} \leftarrow C_{\text{filtered}} \cup \{(x_i, y_i, s_i, m_i)\}$
    end if
end for
return C_{filtered}

Regarding computational complexity, let $N = H \times W$ denote the total pixel count. Feature extraction requires $O(12N)$ operations; ED computation at stride 2 processes $N/4$ pixels with $|D| = 8$ neighbor comparisons per pixel, yielding $O(24N)$; PoI detection and NMS contribute $O(N + C^2)$ where $C \ll N$ is the candidate count after NMS. The overall complexity is therefore $O(N)$, confirming linear scaling with image resolution and supporting real-time deployment on resource-constrained farm hardware.

5.3. Evaluation metrics

System performance is assessed at the pixel level using two complementary metrics. The ROC-AUC (Area Under the Receiver Operating Characteristic Curve) measures the ability of the ED Map anomaly score to rank dead-fish pixels above live-fish and water pixels across all

decision thresholds, regardless of any specific threshold choice. The Average Precision (AP) measures localization quality as the area under the precision–recall curve, penalizing both missed detections and false alarms. Both metrics are computed without any threshold tuning on the test images, reflecting the true unsupervised performance of the proposed framework.

5.4. Scenario 1: High-density pond with non-uniform lighting

The input (800×452 px) is a very high-density pond with non-uniform lighting, shadows, and greenish turbid water; some unusually light-colored fish are suspected dead (see Figure 3a).

The system detected **788 candidates** (764 peak + 24 contour) and produced a panoptic map with **66 instances**: 47 live fish ($\mathcal{C}_t^{\text{live}}$) and 19 dead fish ($\mathcal{C}_t^{\text{dead}}$); remaining pixels assigned to water (*stuff*, $\mathcal{L}$=0). Thresholds $\tau_{\text{high}} = 0.6978$, $\tau_{\text{classify}} = 0.7085$. Dead energy: [0.7093, 0.7499]; live: [0.5761, 0.7063]. Pixel-level: **ROC-AUC** = 0.953, **AP** = 0.420. Results are shown in Figure 3.

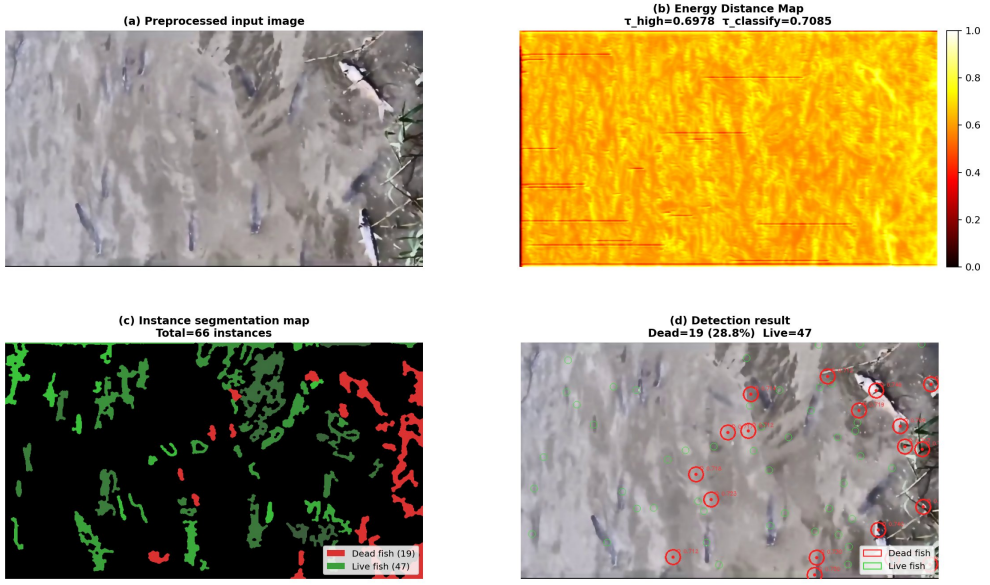

Figure 3: Scenario 1 results: (a) preprocessed input, (b) Energy Distance Map ($\tau_{\text{high}} = 0.6978$), (c) panoptic map (66 instances: 47 live / 19 dead), (d) detection overlay.

5.5. Scenario 2: Medium-density pond with stable lighting

The input (800×452 px) has medium density, uniform lighting, low turbidity, and light-blue water providing good contrast; some fish show slightly atypical coloration (Figure 4a).

The system detected **906 candidates** (870 peak + 36 contour) and produced **90 instances**: 70 live and 20 dead. Thresholds $\tau_{\text{high}} = 0.6043$, $\tau_{\text{classify}} = 0.6050$. Dead energy: [0.6078, 0.6748]; live: [0.4998, 0.6029]. Pixel-level: **ROC-AUC** = 0.914, **AP** = 0.334. Results are shown in Figure 4.

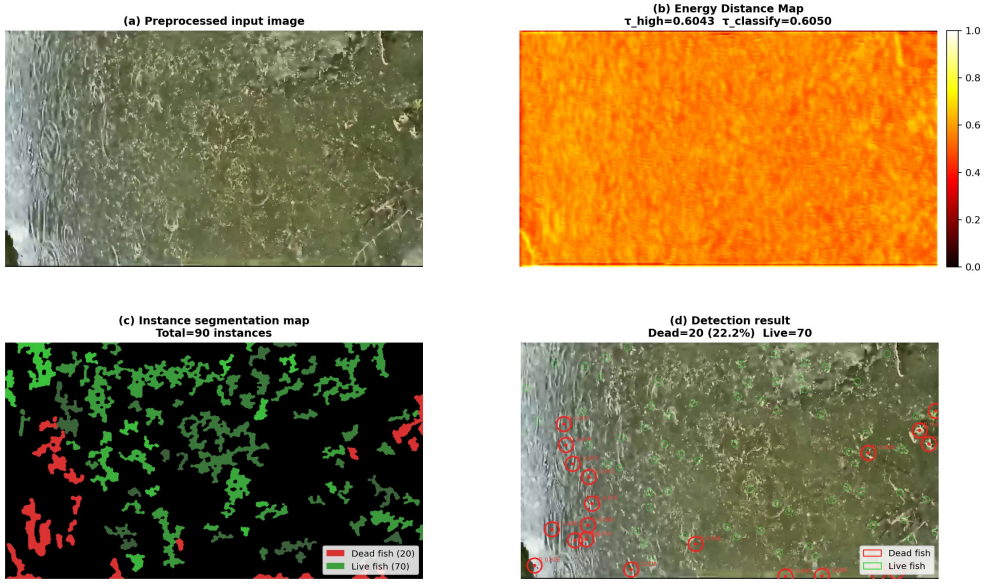

Figure 4: Scenario 2 results: (a) preprocessed input, (b) Energy Distance Map ($\tau_{\text{high}} = 0.6043$), (c) panoptic map (90 instances: 70 live / 20 dead), (d) detection overlay.

5.6. Scenario 3: Low-density pond with high contrast

The input (800×452 px) has the lowest density, best lighting uniformity, and highest water clarity (turquoise), representing near-ideal detection conditions (Figure 5a).

The system detected **758 candidates** (741 peak + 17 contour, 97.8% peak) and produced **56 instances**: 50 live and 6 dead. Thresholds $\tau_{\text{high}} = 0.7442$, $\tau_{\text{classify}} = 0.7653$. Dead energy: [0.7672, 0.8059]; live: [0.5617, 0.7645]. Pixel-level: **ROC-AUC** = 0.855, **AP** = 0.114. Results are shown in Figure 5.

5.7. Summary and discussion

Table 2 summarizes the quantitative results across all three scenarios. Table 3 reports pixel-level metrics.

Table 2: Panoptic segmentation results across three scenarios

Metric	Scenario 1	Scenario 2	Scenario 3
Image size (px)	800×452	800×452	800×452
Candidates detected	788	906	758
Total instances	66	90	56
Dead fish (*things* $\mathcal{C}_t^{\text{dead}}$)	19 (28.8%)	20 (22.2%)	6 (10.7%)
Live fish (*things* $\mathcal{C}_t^{\text{live}}$)	47 (71.2%)	70 (77.8%)	50 (89.3%)
τ_{high} (90th pct)	0.6978	0.6043	0.7442
τ_{classify} (75th pct)	0.7085	0.6050	0.7653

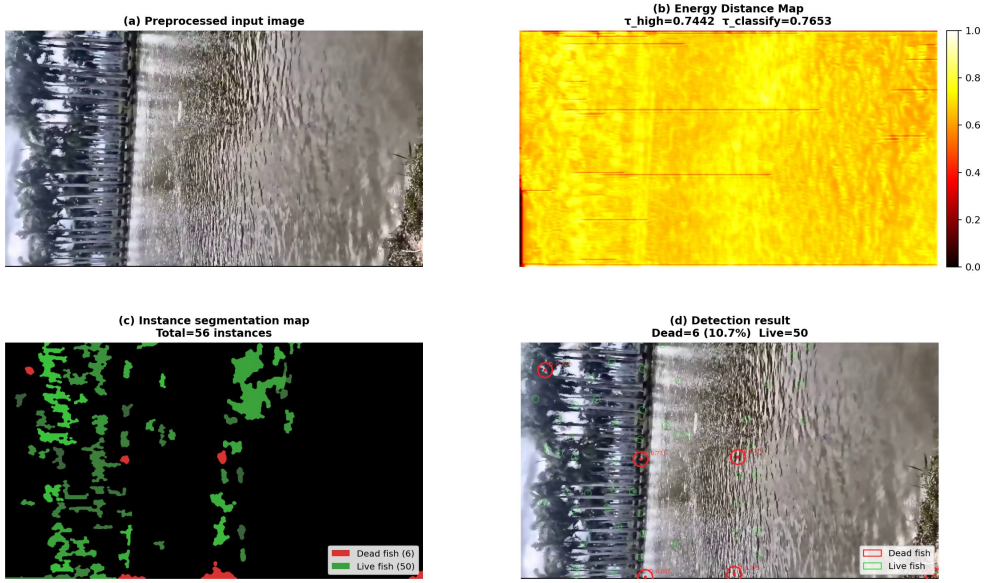

Figure 5: Scenario 3 results: (a) preprocessed input, (b) Energy Distance Map ($\tau_{\text{high}} = 0.7442$), (c) panoptic map (56 instances: 50 live / 6 dead), (d) detection overlay.

Table 3: Pixel-level anomaly detection metrics

Metric	Scenario 1	Scenario 2	Scenario 3
ROC-AUC	**0.9532**	0.9135	0.8545
Average Precision (AP)	**0.4202**	0.3336	0.1142

This study presents an unsupervised approach addressing the critical challenge of scarce labeled mortality data in aquaculture.

Direct comparison with supervised baselines is inherently constrained because no public benchmark dataset exists for dead pangasius detection, and supervised methods such as YOLOv8 or Faster R-CNN require labeled training data unavailable for our experimental pond images. Nevertheless, Table 4 provides a qualitative and contextual comparison of the proposed method against representative deep learning and unsupervised approaches from the literature.

Table 4: Qualitative comparison of the proposed method with representative approaches

Method	Training data	GPU	Domain	Score	Dead fish (AUC)
YOLOv4 [15]	Large labeled	Yes	Specific	BBox	Yes (0.984)
Faster R-CNN [11]	Large labeled	Yes	General	BBox	No
WinCLIP [25]	None	Yes	General	Attention	No
AnoDDPM [26]	Normal only	Yes	General	Recon. err.	No
Proposed	**None**	**No**	**Aquaculture**	**ED Map**	**Yes (≥0.855)**

As shown in Table 4, the proposed method is the only approach that is simultaneously training-free, GPU-free, domain-specific to aquaculture, and directly targeting dead fish detection. The YOLOv4-based method of Zhao et al. [15] achieves ROC-AUC of 0.984 but requires a large labeled dataset under controlled conditions; our method achieves ROC-AUC ≥ 0.855 on real uncontrolled pond imagery with zero annotation cost. WinCLIP and AnoDDPM, while label-free, require GPU inference and are not tailored to dead pangasius signatures.

ROC-AUC decreases from Scenario 1 (0.953) to Scenario 3 (0.855), consistent with decreasing dead-fish count (19, 20, 6), yet all values exceed 0.85, confirming the Energy Distance scoring function reliably separates dead-fish pixel signatures across diverse conditions. The AP decrease in Scenario 3 (0.114 vs. 0.420) reflects the extreme class imbalance at low dead-fish density, a known limitation of AP in sparse positive settings.

6. CONCLUSIONS

This study developed an automatic dead pangasius catfish detection system grounded in an Energy Distance-based Panoptic Segmentation framework, addressing the critical challenge of scarce labeled data. The main achievements are: (1) A formal theoretical model integrating panoptic segmentation formalism, Energy-Based Models, Energy Distance statistics, and a specialized 12-dimensional feature space; (2) A clear functional model separating *stuff* (water background) from *things* (fish instances) via a two-step panoptic strategy; (3) A lightweight CPU-deployable tool with linear computational complexity and no training data requirement; (4) Experimental validation on three real pond scenarios achieving ROC-AUC ≥ 0.855, demonstrating robustness across varying density and lighting.

Future directions include: (1) Multi-type anomaly detection (disease symptoms, abnormal behaviors); (2) Video-based temporal analysis to reduce false positives; (3) Multi-scale feature extraction for fish at varying depths; (4) Integration with IoT environmental sensors (dissolved oxygen, pH, temperature); (5) Long-term field deployment across seasons and growth stages; (6) Semi-supervised extensions leveraging occasional labeled data.

ACKNOWLEDGMENTS

This research was conducted within the framework of the Vietnam National Key Project, Grant No. KC-4.0.41/19-25, "Research and development of digital transformation model applying Industry 4.0 technologies in industrial pangasius catfish farming".

REFERENCES

[1] Food and Agriculture Organization of the United Nations, "The state of world fisheries and aquaculture 2022: Towards blue transformation," FAO, Rome, Tech. Rep., 2022.

[2] S. S. De Silva and N. T. Phuong, "Striped catfish farming in the Mekong Delta, Vietnam: a tumultuous path to a global success story," *Reviews in Aquaculture*, vol. 3, no. 2, pp. 45–73, 2011.

[3] M. Føre, K. Frank, T. Norton, E. Svendsen, J. A. Alfredsen, T. Dempster, H. Eguiraun, W. Watson, A. Stahl, L. M. Sunde, C. Schellewald, K. R. Skøien, M. O. Alver, and D. Berckmans, "Precision

fish farming: A new framework to improve production in aquaculture," *Biosystems Engineering*, vol. 173, pp. 176–193, 2018.

[4] L. T. Phan, T. M. Bui, T. T. T. Nguyen, G. J. Gooley, B. A. Ingram, H. V. Nguyen, P. T. Nguyen, and S. S. De Silva, "Current status of farming practices of striped catfish, Pangasianodon hypophthalmus in the Mekong Delta, Vietnam," *Aquaculture*, vol. 296, no. 3–4, pp. 227–236, 2009.

[5] C. E. Boyd and C. S. Tucker, *Pond Aquaculture Water Quality Management*. Boston, MA: Kluwer Academic Publishers, 1998.

[6] L. Zhang, D. Li, and Q. Wang, "Early detection of fish diseases using computer vision: A review," *Computers and Electronics in Agriculture*, vol. 192, p. 106537, 2022.

[7] X. Yang, S. Zhang, J. Liu, Q. Gao, S. Dong, and C. Zhou, "Deep learning for smart fish farming: applications, opportunities and challenges," *Reviews in Aquaculture*, vol. 13, no. 1, pp. 66–90, 2021.

[8] B. P. Ruff, J. A. Marchant, and A. R. Frost, "Fish sizing and monitoring using a stereo image analysis system applied to fish farming," *Aquacultural Engineering*, vol. 14, no. 2, pp. 155–173, 1995.

[9] M. Saberioon, A. Gholizadeh, P. Cisar, A. Pautsina, and J. Urban, "Application of machine vision systems in aquaculture with emphasis on fish: state-of-the-art and key issues," *Reviews in Aquaculture*, vol. 9, no. 4, pp. 369–387, 2017.

[10] R. J. Petrell, X. Shi, R. K. Ward, A. Naiberg, and C. R. Savage, "Determining fish size and swimming speed in cages and tanks using simple video techniques," *Aquacultural Engineering*, vol. 16, no. 1–2, pp. 63–84, 1997.

[11] A. Salman, S. A. Siddiqui, F. Shafait, A. Mian, M. R. Shortis, K. Khurshid, A. Ulges, and U. Schwanecke, "Automatic fish detection in underwater videos by a deep neural network-based hybrid motion learning system," *ICES Journal of Marine Science*, vol. 77, no. 4, pp. 1295–1307, 2020.

[12] D. Li, Z. Wang, S. Wu, Z. Miao, L. Du, and Y. Duan, "Automatic counting methods in aquaculture: A review," *Journal of the World Aquaculture Society*, vol. 52, no. 2, pp. 269–283, 2021.

[13] W. Wang, Y. Sun, Y. Fang, S. Luo, Y. Dai, and Y. Zhao, "Fish recognition using convolutional neural network and image preprocessing," *Aquacultural Engineering*, vol. 95, p. 102188, 2021.

[14] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[15] A. Salman, A. Jalal, F. Shafait, A. Mian, M. Shortis, J. Seager, and E. Harvey, "Fish species classification in unconstrained underwater environments based on deep learning," *Limnology and Oceanography: Methods*, vol. 14, no. 9, pp. 570–585, 2016.

[16] M.-C. Chuang, J.-N. Hwang, and K. Williams, "A feature learning and object recognition framework for underwater fish images," *IEEE Transactions on Image Processing*, vol. 25, no. 4, pp. 1862–1872, 2016.

[17] C. Zhou, K. Lin, D. Xu, L. Chen, Q. Guo, C. Sun, and X. Yang, "Near infrared computer vision and neuro-fuzzy model-based feeding decision system for fish in aquaculture," *Computers and Electronics in Agriculture*, vol. 146, pp. 114–124, 2018.

[18] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollár, "Panoptic segmentation," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 9404–9413.

[19] A. Kirillov, R. Girshick, K. He, and P. Dollár, "Panoptic feature pyramid networks," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 6399–6408.

[20] B. Cheng, M. D. Collins, Y. Zhu, T. Liu, T. S. Huang, H. Adam, and L.-C. Chen, "Panoptic-DeepLab: A simple, strong, and fast baseline for bottom-up panoptic segmentation," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 12 475–12 485.

[21] R. Mohan and A. Valada, "EfficientPS: Efficient panoptic segmentation," *International Journal of Computer Vision*, vol. 129, pp. 1551–1579, 2021.

[22] G. J. Székely and M. L. Rizzo, "Energy statistics: A class of statistics based on distances," *Journal of Statistical Planning and Inference*, vol. 143, no. 8, pp. 1249–1272, 2013.

[23] M. L. Rizzo and G. J. Székely, "Energy distance," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 8, no. 1, pp. 27–38, 2016.

[24] T. D. Nguyen, T. T. Nguyen, and V. H. Nguyen, "A deep learning approach for real-time fish behavior recognition in aquaculture monitoring systems," *Journal of Computer Science and Cybernetics*, vol. 39, no. 2, pp. 143–160, 2023.

[25] J. Jeong, Y. Zou, T. Kim, D. Zhang, A. Ravichandran, and O. Dabeer, "WinCLIP: Zero-/few-shot anomaly classification and segmentation," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2023, pp. 19 606–19 616.

[26] J. Wyatt, A. Leach, S. M. Schmon, and C. G. Willcocks, "AnoDDPM: Anomaly detection with denoising diffusion probabilistic models using simplex noise," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2022, pp. 650–656.

[27] Y. LeCun, S. Chopra, R. Hadsell, M. A. Ranzato, and F.-J. Huang, "A tutorial on energy-based learning," in *Predicting Structured Data*, G. Bakir, T. Hofmann, B. Schölkopf, A. Smola, and B. Taskar, Eds. Cambridge, MA: MIT Press, 2006. [Online]. Available: http://yann.lecun.com/exdb/publis/pdf/lecun-06.pdf

Received on October 21, 2025
Accepted on May 02, 2026