

DOI: 10.70993/TCDN.2026.3(27).28-37

MÔ HÌNH YOLOV11 CẢI TIẾN VỚI KHỐI KẾT HỢP KHÔNG GIAN VÀ KÊNH TRONG PHÁT HIỆN BỆNH CÁ

Đoàn Hồng Quang¹, Đặng Tiến Quyền², Hà Hải Nam³,
Nguyễn Huy Hưng⁴, Lê Thị Thu⁵

Ngày nhận bài: 25/11/2025; Gửi phản biện: 12/12/2025; Duyệt đăng: 15/01/2026

Tóm tắt: Nghiên cứu này đề xuất một phương pháp cải tiến dựa trên mô hình YOLOv11 cho bài toán phát hiện bệnh cá trong môi trường nuôi trồng thủy sản. Các cải tiến chính bao gồm: (1) tích hợp khối kết hợp đặc trưng không gian và kênh (Spatial and Channel Fusion Block - SCFB) để tăng cường khả năng trích xuất đặc trưng; (2) sử dụng hàm mất mát Scylla-IoU (SIoU) để tối ưu hóa quá trình huấn luyện; và (3) bổ sung tầng đầu ra P2 chuyên biệt cho việc phát hiện các vật thể có kích thước nhỏ. Mô hình được đánh giá trên bộ dữ liệu FISH-DISEASE-DT (12.556 ảnh, 5 loài cá và 4 loại bệnh). Kết quả cho thấy mô hình đề xuất đạt độ chính xác trung bình (mAP@0.5-0.95) là 84,95%, vượt trội so với các mô hình nền tảng như SSD (47,3%), Faster R-CNN (55,62%), DETR (64,82%) và YOLOv11 gốc (71,67%). Về hiệu suất, mô hình đạt tốc độ xử lý 89 khung hình/giây (Frames Per Second - FPS) trên ảnh đầu vào 640×640 pixel, đáp ứng yêu cầu ứng dụng thời gian thực. Kết quả này khẳng định tiềm năng của phương pháp trong việc phát triển hệ thống hỗ trợ chẩn đoán sớm, góp phần nâng cao hiệu quả quản lý và tính bền vững trong ngành thủy sản.

Từ khoá: YOLOv11, Kết hợp đặc trưng không gian-kênh, Thị giác máy tính trong nông nghiệp, Phát hiện bệnh tự động, Sức khỏe động vật thủy sản.

Abstract: This study proposes an enhanced method for fish disease detection in aquaculture, based on an improved YOLOv11 model. Key enhancements include: (1) integrating a SCFB to strengthen feature representation; (2) employing the Scylla-IoU (SIoU) loss function for optimized training convergence; and (3) adding a dedicated P2 detection head to improve small-object detection sensitivity. The model was evaluated on the FISH-DISEASE-DT dataset (12,556 images covering 5 fish species and 4 disease types). The proposed model achieved a mean Average Precision (mAP@0.5-0.95) of 84.95%, outperforming baseline models such as SSD (47.3%), Faster R-CNN (55.62%), DETR (64.82%), and the original YOLOv11 (71.67%). In terms of efficiency, it maintains an average processing speed of 89 FPS on 640×640 pixel inputs, meeting real-time application requirements. These findings confirm the potential of the proposed method for developing early warning systems, contributing to more effective and sustainable aquaculture management.

Keywords: YOLOv11, Spatial-Channel Feature Fusion, Computer Vision in Agriculture, Automated Disease Detection, Aquatic Animal Health.

^{1,4} Viện Ứng dụng công nghệ

^{2,3} Trường Đại học Điện lực.

⁵ Học viện Chính trị quốc gia Hồ Chí Minh

1. Đặt vấn đề

Phát hiện sớm bệnh cá là yếu tố then chốt trong quản lý nuôi trồng thủy sản (FAO, 2024). Tuy nhiên, các mô hình hiện tại chủ yếu tập trung vào nhận diện loài hoặc bệnh lý khi đã đưa lên bờ, trong phòng lab, hoặc trên phông nền sạch, chưa tối ưu cho môi trường dưới nước do thách thức về ánh sáng, độ đục và chuyển động tự nhiên (Wang et al., 2023; Yang et al., 2024). Các phương pháp truyền thống như SSD (Liu et al., 2016) (47,3% mAP), Faster R-CNN (Ren et al., 2016) (55,62%) và DETR (Carion et al., 2020) (64,82%) bộc lộ hạn chế trong việc xử lý vật thể đa kích thước, đặc biệt là hiện tượng suy giảm đặc trưng hình học của đối tượng nhỏ qua các lớp tích chập. Ngay cả YOLOv11 (Glenn Jocher & Jing Qiu, 2024) (71,64%), dù đã cải thiện khả năng tổng hợp đa tỉ lệ, vẫn gặp khó khăn khi đối diện với các đặc trưng hình thái bệnh lý đa dạng. Trong khi các nghiên cứu cải tiến gần đây, chẳng hạn như FishDet-YOLO (Yang et al., 2024), chủ yếu tập trung vào việc mô hình hóa các phụ thuộc tầm xa (long-range dependencies), và các phương pháp khác chỉ giải quyết từng phần các hạn chế như khả năng phát hiện vật thể nhỏ hoặc tối ưu hóa hàm mất mát, nghiên cứu này đề xuất một hướng tiếp cận toàn diện và khác biệt. Cụ thể, chúng tôi đồng thời giải quyết ba khía cạnh then chốt: (1) tăng cường biểu diễn đặc trưng thông qua cơ chế hợp nhất không gian-kênh, (2) tối ưu hóa độ chính xác định vị, và (3) nâng cao độ nhạy phát hiện đối với các vật thể kích thước nhỏ.

Nghiên cứu này đề xuất mô hình YOLO-SCFB được cải tiến từ YOLOv11 thông qua việc tích hợp: (i) Mô-đun SCFB: Kiến trúc đa nhánh kết hợp không gian và kênh; (ii) Hàm mất mát SIOU: Tăng cường độ chính xác định vị (Gevorgyan, 2022); (iii) Tầng Head P2: Tối ưu việc trích xuất đặc trưng cho đối tượng nhỏ (Zhai et al., 2023).

Mô hình được thực nghiệm trên bộ dữ liệu FISH-DISEASE-DT (Dang, 2025) để chứng minh hiệu suất vượt trội so với các phương pháp truyền thống trong việc phát hiện bệnh lý của cá trực tiếp dưới nước. Câu hỏi nghiên cứu: Mô hình YOLO-SCFB có cải thiện hiệu suất so với các phương pháp truyền thống không?

2. Phương pháp nghiên cứu

2.1. Tổng quan phương pháp

Nghiên cứu xây dựng mô hình YOLO-SCFB dựa trên kiến trúc YOLOv11, tập trung tối ưu khả năng trích xuất và duy trì đặc trưng bệnh lý. Giải pháp bao gồm việc tích hợp khối SCFB để khai thác mối tương quan giữa không gian và kênh, kết hợp cùng tầng Head P2 nhằm tăng cường độ phân giải cho các đối tượng kích thước nhỏ. Hàm kích hoạt SiLU được lựa chọn nhằm tận dụng đặc tính có đạo hàm liên tục, qua đó giảm thiểu hiện tượng tiêu biến gradient và cải thiện luồng lan truyền ngược so với hàm ReLU truyền thống, một yếu tố quan trọng để ổn định quá trình huấn luyện cho các kiến trúc mạng có độ sâu trung bình. Về hàm mất mát, nghiên cứu áp dụng SIOU để tối ưu hóa đồng thời tốc độ hội tụ và độ chính xác định vị hộp giới hạn. SIOU được ưu tiên vì cơ chế toàn diện của nó, không chỉ đánh giá độ chồng lấp hình học (IoU) mà còn tích hợp các thành phần phạt cho góc lệch phương vị và khoảng cách giữa các tâm. Cơ chế này định hướng quá trình tối ưu một cách hiệu quả hơn, dẫn đến sự hội tụ nhanh chóng và ổn định vượt trội so với các biến thể hàm mất mát dựa trên IoU thông thường như CIOU hay DIOU. Sự phối hợp giữa các thành phần này giúp giảm thiểu việc mất mát thông tin ở các lớp mạng sâu, đảm bảo hiệu quả phát hiện cá và bệnh của cá trong môi trường dưới nước, nhiều.

2.2. Dữ liệu và kỹ thuật tăng cường

Nghiên cứu sử dụng bộ dữ liệu FISH-DISEASE-DT, bao gồm 12,556 ảnh với 19,770 nhãn thuộc 9 lớp đối tượng (5 loài cá và 4 loại bệnh). Dữ liệu được phân chia theo tỷ lệ: 70% cho huấn luyện, 20% xác thực và 10% kiểm tra (Bảng 1).

Bảng 1. Mô tả bộ dữ liệu cho mô hình

Tên lớp	Số lượng mẫu	Mẫu dữ liệu		
		Huấn luyện	Xác thực	Kiểm tra
Bighead	1.925	1,379	353	193
Carp	2.282	1,558	476	248
EUS	3.591	2,599	658	334
Eye disease	2.137	1515	426	196
Fin lesions	2.367	1,627	483	257
Grass Carp	1.899	1375	358	166
Rotten Gills	2.003	1,417	390	196
Snakehead	1.860	1,260	418	182
Tilapia	1.706	1,180	362	164
Tổng cộng	19.770	13.910	3.924	1.936

Quy trình tăng cường dữ liệu:

1. Biến đổi hình học: Áp dụng lật ngang và dọc, xoay 90° , xoay ngẫu nhiên $\pm 15\%$ và biến dạng khung hình $\pm 10\%$ để đa dạng hóa góc chụp thực tế.

2. Hiệu chỉnh độ bão hòa $\pm 10\%$ và thêm 0.1% nhiễu kỹ thuật số nhằm tăng độ bền vững của mô hình trước điều kiện ánh sáng và chất lượng ảnh biến động.

3. Đồng nhất kích thước toàn bộ dữ liệu về 640×640 pixel để đảm bảo hiệu suất xử lý ổn định.

Tăng cường dữ liệu giúp mô hình nâng cao khả năng tổng quát h36

óa và tránh quá khớp. Việc tích hợp các biến đổi hình học và nhiễu kỹ thuật giúp thuật toán bền vững hơn trước các sai số như ánh sáng hay góc chụp. Đồng thời, chuẩn hóa đầu vào giúp cân bằng giữa độ chính xác và tốc độ, giúp mô hình hoạt động hiệu quả ngay cả trong môi trường thực tế phức tạp.

2.3. Kiến trúc Spatial-Channel Fusion Block

Để tăng khả năng trích xuất đặc trưng, đặc biệt đối với sự phức tạp của các đặc trưng vùng bệnh có hình dạng, kích thước, màu sắc đa dạng, mô hình được mở rộng bằng việc tích hợp SCFB là một khối đa nhánh, kết hợp thông tin không gian và kênh để tăng cường biểu diễn của bản đồ đặc trưng (Hình 1).

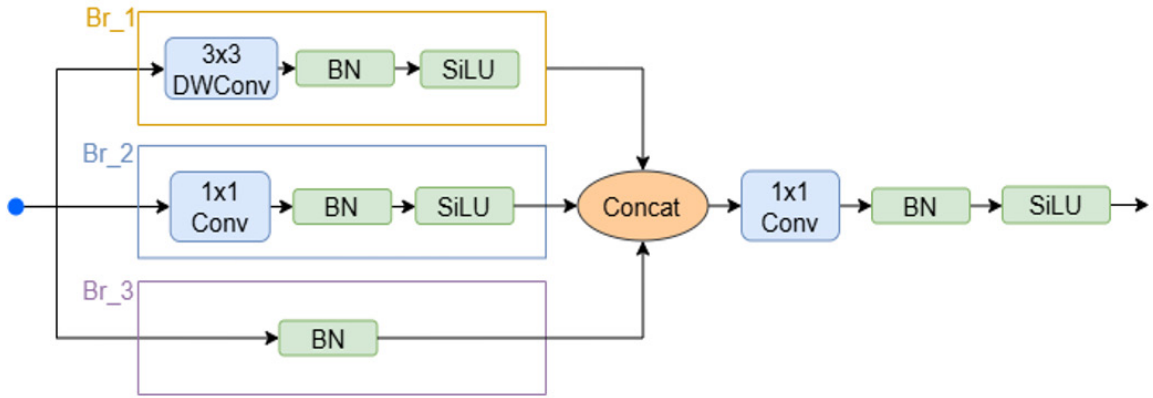
Kiến trúc SCFB bao gồm 3 nhánh chính và một nhánh hợp nhất đầu ra: Nhánh không gian (*Branch 1*) sử dụng tích chập theo chiều kênh với ma trận tích chập 3×3 để trích xuất đặc trưng không gian, giúp duy trì chi tiết hình học (Chollet, 2016). Cho đặc trưng đầu vào $X \in \mathbb{R}^{B \times C \times H \times W}$, đầu ra của nhánh này tính theo công thức (1).

$$X_s = SiLU \left(BN(DWConv_{3 \times 3}(X)) \right) \quad (1)$$

Trong đó:

- $DWConv_{3 \times 3}$: tích chập theo chiều kênh, kích thước ma trận tích chập 3×3 ,
- BN : chuẩn hóa theo lô (Batch Normalization),

Hình 1. Kiến trúc khối Spatial-Channel Fusion Block



Nhánh kênh (Branch 2) sử dụng tích chập điểm với ma trận tích chập 1×1 để học mối quan hệ giữa các kênh (Lin et al., 2013), bổ sung thông tin nhánh không gian có thể bỏ qua:

$$X_c = SiLU \left(BN \left(Conv_{1 \times 1}(X) \right) \right) \quad (2)$$

Nhánh ổn định (Branch 3) đầu vào được đưa qua lớp chuẩn hóa theo lô, có chức năng chuẩn hóa giá trị của từng kênh theo phân phối trung bình và phương sai của lô hiện tại. Quá trình này giúp ổn định miền giá trị của đặc trưng, từ đó giảm thiểu hiện tượng dịch chuyển phân phối nội bộ, hỗ trợ quá trình huấn luyện hội tụ nhanh và ổn định hơn.

$$X_b = BN(X) \quad (3)$$

Hợp nhất và tinh chỉnh đầu ra của ba nhánh trước được hợp nhất theo chiều kênh, sau đó đi qua tích chập điểm 1×1 để hợp nhất và tinh chỉnh đặc trưng:

$$Y = SiLU \left(BN \left(Conv_{1 \times 1} \left(Concat[X_b, X_s, X_c] \right) \right) \right) \quad (4)$$

SCFB được biểu diễn dưới dạng một hàm ánh xạ phi tuyến từ không gian đặc trưng đầu vào sang không gian đặc trưng đầu ra có cùng kích thước:

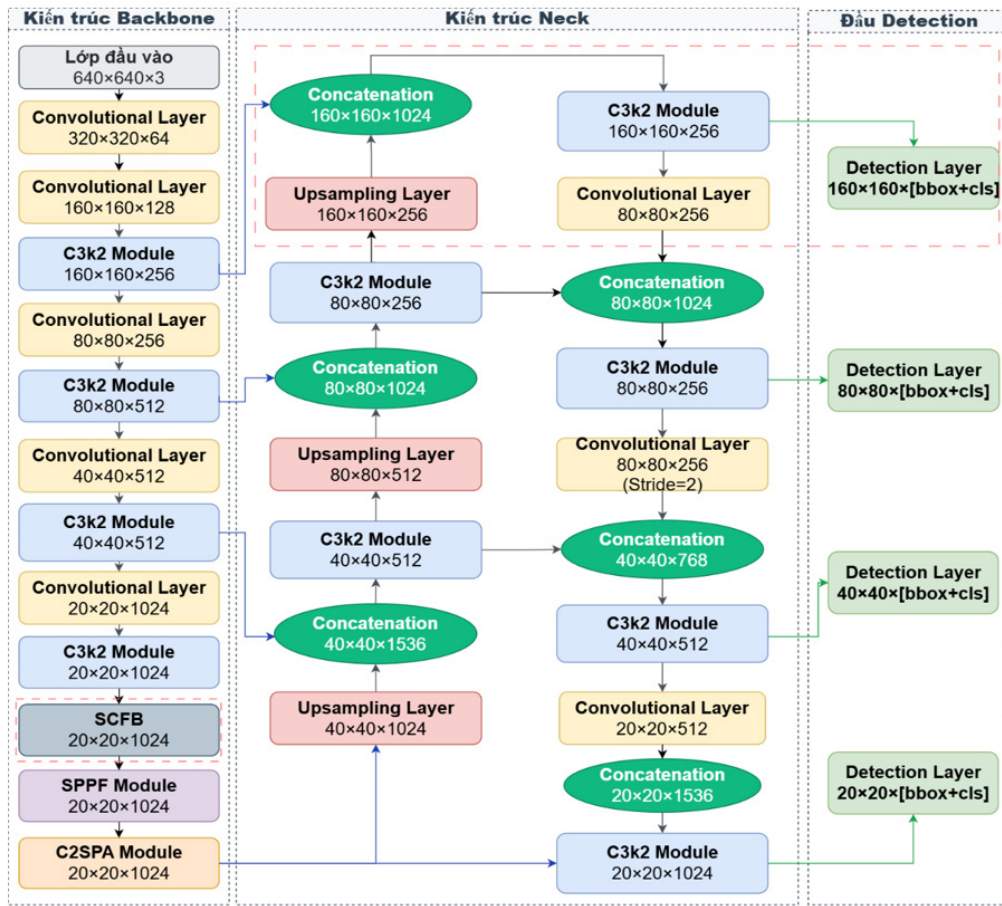
$$f_{SCFB}: \mathbb{R}^{B \times C \times H \times W} \rightarrow \mathbb{R}^{B \times C \times H \times W}, Y = f_{SCFB}(X) \quad (5)$$

SCFB xử lý đồng thời thông tin không gian và kênh thông qua ba nhánh tương tác bổ trợ. Nhánh không gian tập trung bảo toàn chi tiết hình học, nhánh kênh tăng cường khả năng biểu diễn thông tin giữa các kênh. Quá trình hợp nhất tổng hợp đặc trưng từ cả hai khía cạnh, đảm bảo bản đồ đặc trưng đầu ra giữ nguyên kích thước $H \times W$ số kênh C . Nhờ đó, SCFB tăng khả năng mô hình hóa các đặc trưng từ các vùng bệnh, cải thiện hiệu suất phát hiện và nâng cao độ chính xác của mô hình.

2.4. Mô hình cải tiến

Khối SCFB được đặt tại đầu ra P5 (trước SPPF) để tổng hợp và nâng cao đặc trưng từ Backbone, qua đó kết hợp thông tin ngữ cảnh tổng thể với các đặc điểm cục bộ.

Hình 2. Kiến trúc mô hình YOLO-SCFB



Kiến trúc đề xuất tối ưu hóa mối quan hệ không gian-kênh mà không làm gián đoạn trường tiếp nhận đa tỷ lệ, cung cấp các biểu diễn đặc trưng nhiều thông tin cho giai đoạn Neck để thực hiện việc kết hợp đặc trưng đa cấp độ một cách hiệu quả. Việc bổ sung đầu ra P2 tăng cường độ nhạy trong việc phát hiện các đối tượng có kích thước nhỏ, trong khi vẫn duy trì độ chính xác ổn định đối với các đối tượng ở tỷ lệ lớn hơn. Hiệu suất định vị được cải thiện rõ rệt thông qua hàm mất mát SIOU, vốn tối ưu hóa đồng thời các tham số vị trí và hình học của hộp giới hạn trong quá trình huấn luyện. Tổng thể, các cải tiến này hoạt động một cách hiệp đồng, tối ưu hóa luồng lan truyền đặc trưng và nâng cao độ tin cậy của mô hình trong các điều kiện môi trường thực tế đa dạng và phức tạp.

3. Kết quả và thảo luận

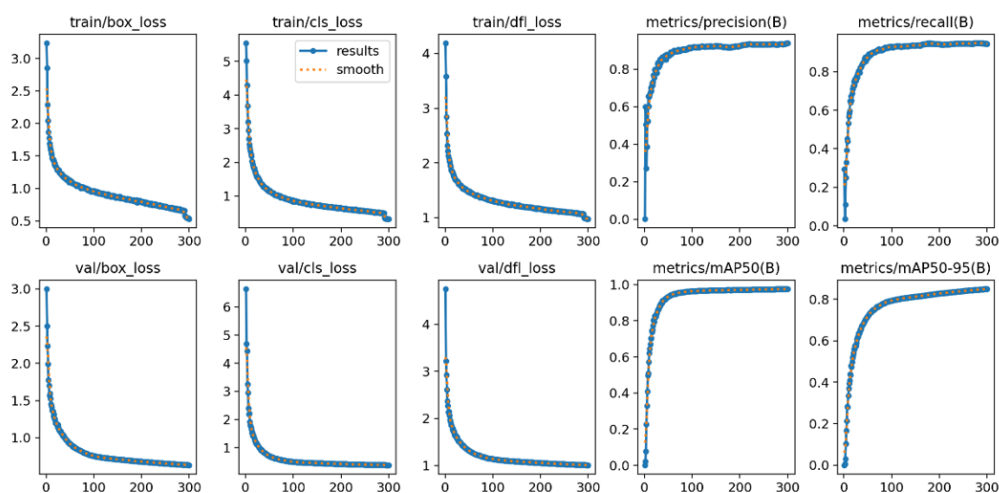
Quá trình huấn luyện mô hình được triển khai trên nền tảng Kaggle Notebook, sử dụng môi trường điện toán đám mây với cấu hình phần cứng mặc định gồm hai GPU NVIDIA T4 (16 GB VRAM mỗi GPU), CPU 8 nhân, và RAM hệ thống là 30 GB (Harisudhan S, 2025; Kaggle Documentation, n.d.). Việc sử dụng hệ thống này giúp tận dụng khả năng tính toán song song, đảm bảo tốc độ huấn luyện cao và khả năng lưu trữ mô hình trung gian ổn định.

Môi trường phần mềm được thiết lập thống nhất nhằm đảm bảo tính tái lập kết quả. Các thành phần chính bao gồm Python 3.11.13, PyTorch 2.6.0+cu124.

3.1. Kết quả huấn luyện

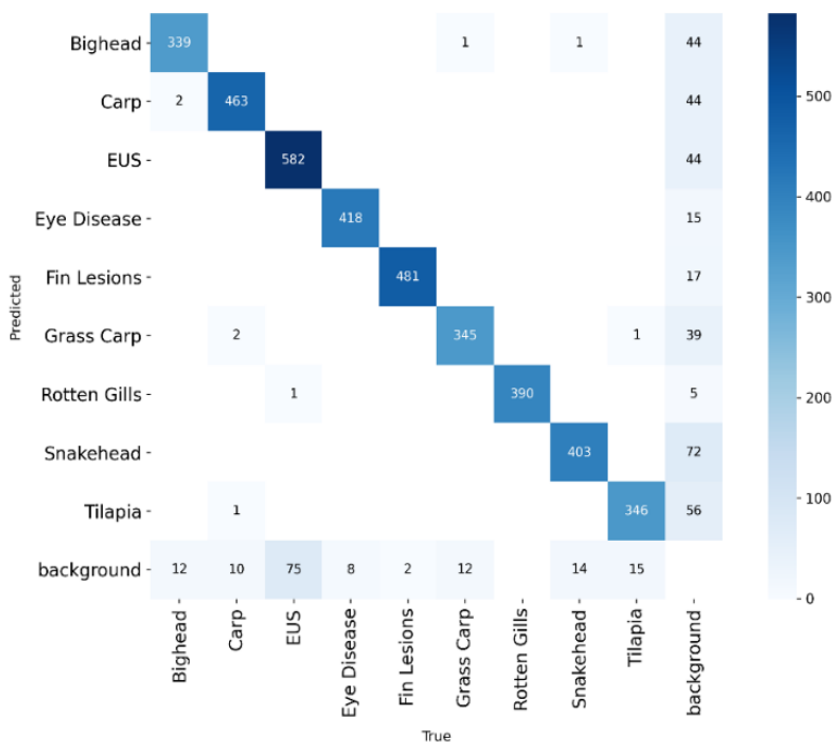
Hình 3 cho thấy YOLO-SCFB hội tụ ổn định với hàm mất mát giảm nhanh ở giai đoạn đầu và duy trì xu hướng giảm đến cuối huấn luyện. Precision và Recall chỉ dao động nhẹ trước khi đạt cân bằng. Giá trị mAP@0.5 hơn 97% và mAP@0.5:0.95 gần 85%, chứng tỏ mô hình đạt độ chính xác cao và ổn định trên nhiều ngưỡng IoU.

Hình 3. Biểu đồ quá trình huấn luyện mô hình YOLO-SCFB



Bảng 2 phân tích mức đóng góp của từng thành phần cải tiến trên tập dữ liệu kiểm tra. Qua phân tích mô hình khi tích hợp cả ba thành phần (Head P2 + SiLU + SCFB) đạt hiệu suất cao nhất với những cải thiện đáng kể so với phiên bản YOLOv11: mAP@0.5 tăng 3,19%, mAP@0.5-0.95 tăng 13,28%, Precision tăng 3,6% và Recall tăng 6,34%.

Hình 4. Ma trận nhầm lẫn của mô hình YOLO-SCFB



Bảng 2. Ảnh hưởng của các thành phần cải tiến đến hiệu suất mô hình

Thành phần tích hợp	mAP@0.5 (%)	mAP@0.5-0.95 (%)	Precision (%)	Recall (%)
YOLOv11	94,39	71,67	90,31	88,12
+ Head P2	94,61	72,07	89,95	88,96
+ SIoU	95,02	73,18	91,47	88,75
+ SCFB	96,63	83,68	93,51	93,28
Head P2 + SIoU +SCFB	97,58	84,95	93,91	94,46

So sánh hiệu suất của YOLO-SCFB với các mô hình khác nhằm chứng minh hiệu quả của mô hình (Bảng 3). Tuy giảm 14 FPS so với bản gốc (103 FPS) nhưng vẫn đạt 89FPS với hiệu suất vượt trội, mô hình vẫn đảm bảo khả năng xử lý thời gian thực.

Bảng 3. So sánh hiệu suất YOLO-SCFB

Model	mAP@0.5 (%)	mAP@0.5-0.95 (%)	Precision (%)	Recall (%)	FPS
SSD	75,03	47,30	68,50	78,20	22
Faster R-CNN	81,28	55,62	74,78	82,55	2
DETR	86,34	64,82	88,41	80,03	20
YOLOv11	94,39	71,67	90,31	88,12	103
YOLO-SCFB	97,58	84,95	93,91	94,46	89

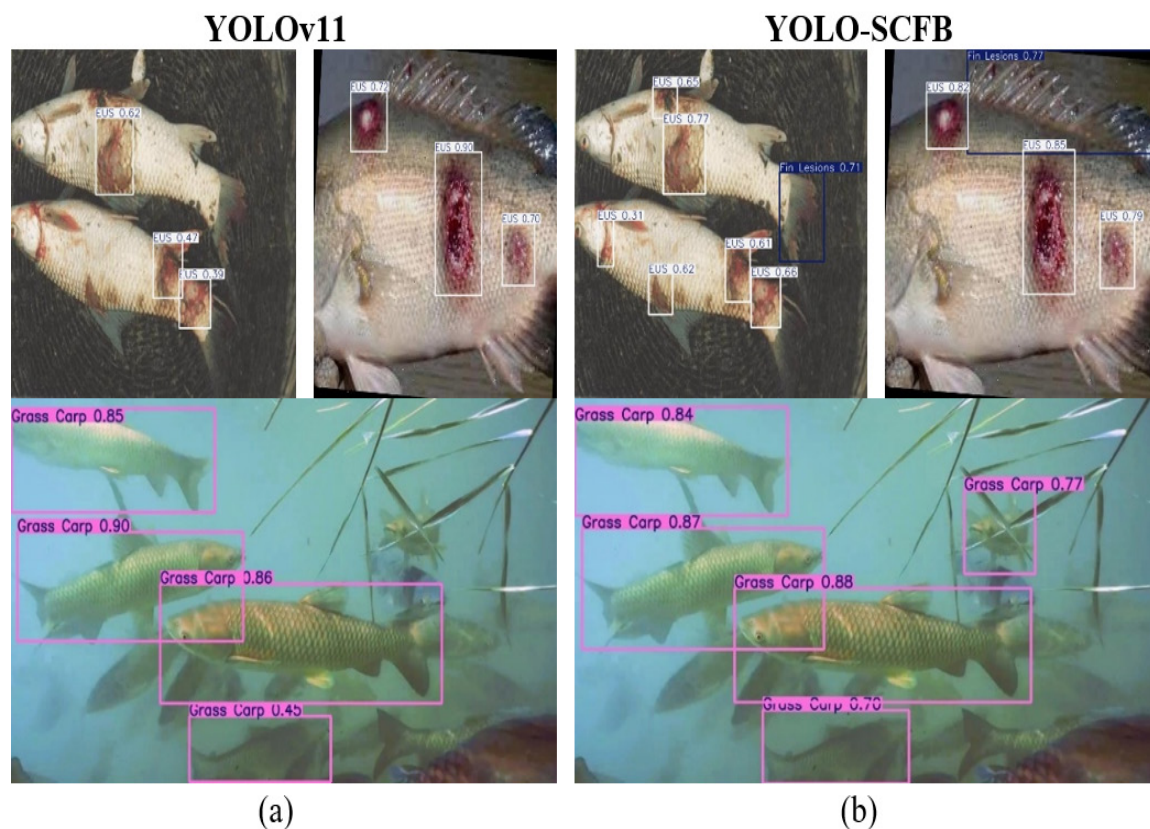
Mô hình đạt 3.767 phát hiện chính xác. Tuy nhiên, vẫn tồn tại các lỗi với phân bố như sau: lỗi gán nhầm nhãn giữa các lớp (9 trường hợp), lỗi phát hiện giả từ vùng nền (336 trường hợp), và lỗi bỏ sót đối tượng (148 trường hợp).

3.2. Thảo luận

Mô hình YOLO-SCFB đạt hiệu suất vượt trội với $mAP@0.5-0.95$ là 84,95%, so với các mô hình SSD (47,30%), Faster R-CNN (55,62%), DETR (64,82%) và YOLOv11 gốc (71,67%), với tốc độ xử lý 89 FPS đáp ứng yêu cầu thời gian thực.

Ma trận nhầm lẫn (Hình 4), sai số chủ yếu đến từ việc nhận diện nhầm nền thành đối tượng (7,88%), tập trung ở Snakehead và Tilapia do nhiễu từ phản sáng hoặc nước đục. Tỷ lệ bỏ sót và nhầm nhãn thấp (3,77% và 0,22%). Lớp Rotten Gills đạt kết quả tối ưu tốt, lớp EUS bỏ sót cao nhất (75 mẫu) do vùng bệnh nhỏ và biên mờ.

Hình 5. So sánh kết quả phát hiện giữa mô hình YOLOv11 và YOLO-SCFB



Hình 5 minh họa kết quả phát hiện của mô hình gốc YOLOv11 (a) và mô hình đề xuất YOLO-SCFB (b). Hình (a) cho thấy YOLOv11 phát hiện 6 vùng tổn thương EUS với điểm tin cậy từ 0.39 đến 0.90, và 4 cá thể Grass Carp với điểm tin cậy từ 0.45 đến 0.90. Trong khi đó, hình (b) YOLO-SCFB không chỉ duy trì khả năng phát hiện các đối tượng này mà còn cải thiện đáng kể: (1) phát hiện thêm 2 tổn thương Fin Lesions (0.71, 0.77) mà YOLOv11 đã bỏ sót, (2) nâng cao điểm tin cậy trung bình cho các phát hiện Grass Carp (đặc biệt tăng từ 0.45 lên 0.70 cho cá thể khó), và (3) duy trì độ tin cậy ổn định cho các vùng EUS. Kết quả này minh chứng cho khả năng vượt trội của YOLO-SCFB trong việc giảm thiểu lỗi bỏ sót, cải thiện độ tin cậy và mở rộng phạm vi phát hiện so với mô hình gốc.

Hiệu quả của mô hình YOLO-SCFB (Bảng 2), mức tăng 13,28% ($mAP@0.5-0.95$) là kết quả của tính tương hỗ chặt chẽ: Head P2 tăng cường khả năng phát hiện các đặc trưng nhỏ (khắc phục

hạn chế đối với EUS), SCFB và hàm mất mát SIOU tối ưu trích xuất đặc trưng không gian giúp phân biệt rõ các đối tượng có hình thái tương đồng. Tuy tốc độ suy luận giảm 14 FPS so với YOLOv11, nhưng đây là sự đánh đổi khả quan để đạt độ chính xác cao. Kết quả thực nghiệm chứng minh YOLO-SCFB cân bằng tốt giữa tốc độ và độ chính xác, khả thi trong các hệ thống giám sát nuôi trồng thủy sản thông minh.

4. Kết luận

Nghiên cứu này tập trung cải tiến kiến trúc YOLOv11 nhằm giải quyết bài toán phát hiện bệnh sớm trên cá trong môi trường nuôi trồng thủy sản. Thông qua việc sử dụng khối SCFB, bổ sung tầng phát hiện vật thể nhỏ (Head P2), sử dụng hàm mất mát SIOU. Mô hình đề xuất đạt được những cải thiện về độ chính xác và tính toán ổn định, đảm bảo khả thi cho các ứng dụng giám sát thời gian thực.

Kết quả thực nghiệm cho thấy, mô hình YOLO-SCFB đề xuất đạt độ chính xác trung bình (mAP@0.5-0.95) là 84,95%, tăng 13,28% so với phiên bản gốc YOLOv11 (71,67%). Mặc dù độ phức tạp tính toán tăng, nhưng mô hình vẫn duy trì được tốc độ xử lý 89 FPS, vượt trội so với các kiến trúc hiện đại (SSD, Faster R-CNN và DETR).

Dù đạt kết quả tích cực, mô hình vẫn cần được cải thiện trong môi trường phức tạp, nhiễu nhiễu (nước đục, ánh sáng thay đổi). Hướng nghiên cứu tiếp theo sẽ củng cố dữ liệu và tối ưu hóa mô hình cho các thiết bị phần cứng có tài nguyên hạn chế, tăng khả năng ứng dụng của mô hình vào thực tiễn trong ngành nuôi trồng thủy sản.

TÀI LIỆU THAM KHẢO

1. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020, August). End-to-end object detection with transformers. In European conference on computer vision (pp. 213-229). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-58452-8_13.
2. Chollet, F. (2017). *Xception: Deep learning with depthwise separable convolutions*. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1251-1258). <https://doi.org/10.1109/CVPR.2017.195>
3. Dang, Q. (2025, June 23). *FISH-DISEASES-DT*. Roboflow. Retrieved June 23, 2025, from <https://universe.roboflow.com/fish-v4/fish-diseases-dt>
4. FAO. (2024). *The State of World Fisheries and Aquaculture 2024*. In FAO (Ed.), <https://doi.org/10.4060/CD0683EN>
5. Gevorgyan, Z. (2022). *SIOU loss: More powerful learning for bounding box regression*. arXiv preprint arXiv:2205.12740.
6. Glenn Jocher, & Jing Qiu. (2024). Ultralytics YOLO11 [Computer software]. Retrieved from <https://github.com/ultralytics/ultralytics>
7. Harisudhan S. (2025, February 11). *Multi-GPU training in Kaggle: Harnessing 2 × T4 for faster model training*. Medium. Retrieved from <https://medium.com/@speaktoharisudhan/multi-gpu-training-in-kaggle-harnessing-2-t4-for-faster-model-training-tensorflow-f091c22db99e>
8. Kaggle Documentation. (n.d.). *Efficient GPU Usage*. Retrieved January 10, 2025, from <https://www.kaggle.com/docs/efficient-gpu-usage>
9. Lin, M., Chen, Q., & Yan, S. (2013). *Network in network*. arXiv preprint arXiv:1312.4400.
10. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016, September). Ssd: Single shot multibox detector. In European conference on computer vision (pp. 21-37). Cham:

Springer International Publishing. https://doi.org/10.1007/978-3-319-46448-0_2

11. Ren, S., He, K., Girshick, R., & Sun, J. (2016). *Faster R-CNN: Towards real-time object detection with region proposal networks*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), 1137–1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
12. Wang, Z., Liu, H., Zhang, G., Yang, X., Wen, L., & Zhao, W. (2023). *Diseased Fish Detection in the Underwater Environment Using an Improved YOLOV5 Network for Intensive Aquaculture, Fishes*, 8 (2023), 169. <https://doi.org/10.3390/FISHES8030169>
13. Yang, C., Xiang, J., Li, X., & Xie, Y. (2024). *FishDet-YOLO: Enhanced Underwater Fish Detection with Richer Gradient Flow and Long-Range Dependency Capture through Mamba-C2f*. Electronics, 13(18), 3780. <https://doi.org/10.3390/ELECTRONICS13183780>
14. Zhai, X., Huang, Z., Li, T., Liu, H., & Wang, S. (2023). *YOLO-Drone: An Optimized YOLOv8 Network for Tiny UAV Object Detection*. Electronics, 12(17), 3664. <https://doi.org/10.3390/ELECTRONICS12173664>