

PHÁT HIỆN TÀU THUYỀN TỪ KHÔNG ẢNH SỬ DỤNG KIẾN TRÚC YOLO26M: PHÂN TÍCH ĐỊNH LƯỢNG VÀ THỰC NGHIỆM

SHIP DETECTION FROM AERIAL IMAGERY USING YOLO26M ARCHITECTURE: QUANTITATIVE ANALYSIS AND EXPERIMENTAL RESULTS

PHAN THANH NGHI^{1,*}, NGUYỄN MẠNH CƯỜNG¹, TRẦN ĐÌNH LÂN²

¹Trường Đại học Hàng hải Việt Nam

²Viện Khoa học Công nghệ Năng lượng và Môi trường

*Email liên hệ: nghi.khhh2211d@ps.vimaru.edu.vn

DOI: <https://doi.org/10.65154/jmst.994>

Tóm tắt

Phát hiện tàu thuyền giữ vai trò then chốt trong giám sát hàng hải và an ninh vùng biển. Tuy nhiên, đặc thù của ảnh không gian là các đối tượng thường có kích thước nhỏ (chiều cao chỉ chiếm 0.05-0.1 khung hình) và chịu nhiễu lớn từ bề mặt nước. Bài báo này trình bày phương pháp ứng dụng kiến trúc YOLO26m kết hợp với huấn luyện chuyển tiếp để giải quyết bài toán trên. Mô hình được huấn luyện trên bộ dữ liệu "Kaggle Ships in Google Earth" gồm 1293 ảnh, sử dụng GPU Tesla P100. Kết quả thực nghiệm sau 70 epoch cho thấy phương pháp đề xuất có hiệu năng cao với $mAP@0.5$ đạt 0,971 và $mAP@0.5:0.5$ đạt 0.766. Bên cạnh đó, mô hình duy trì sự cân bằng tốt giữa Precision (0,96) và Recall (0,92), đồng thời đạt tốc độ suy luận trung bình 14,1 ms/ảnh (tương đương 58,8FPS). Những kết quả này khẳng định khả năng của YOLO26m trong việc đáp ứng các yêu cầu xử lý thời gian thực cho hệ thống giám sát tự động.

Từ khóa: YOLO26m, phát hiện tàu, không ảnh, giám sát hàng hải.

Abstract

Ship detection plays a crucial role in maritime surveillance and coastal security. However, aerial imagery presents significant challenges due to the small size of objects (typically occupying 0.05-0.1 of the frame height) and interference from water surface noise. This paper presents a method applying the YOLO26m architecture combined with transfer learning to address this task. The model was trained on the "Kaggle Ships in Google Earth" dataset comprising 1293 images, utilizing a Tesla P100 GPU. Experimental results after 70 epochs demonstrate superior performance, with a

$mAP@0.5$ of 0.971 and $mAP@0.5:0.95$ of 0.766. Furthermore, the model maintains a strong balance between Precision (0.96) and Recall (0.92), while achieving an average inference speed of 14.1 ms/image (approximately 58.8 FPS). These results confirm the capability of YOLO26m to meet real-time processing requirements for automated surveillance systems.

Keywords: YOLO26m, Ship Detection, Aerial Imagery, Maritime Surveillance.

1. Mở đầu

Giám sát hàng hải đóng vai trò then chốt trong an ninh quốc phòng, quản lý hàng hải. Với sự phát triển của công nghệ viễn thám, việc phát hiện tàu thuyền từ ảnh vệ tinh quang học và không ảnh đã trở thành một bài toán cấp thiết nhằm tự động hóa quy trình giám sát diện rộng [1], [2]. Tuy nhiên, đây là một thách thức lớn trong lĩnh vực trí tuệ nhân tạo do đặc thù của đối tượng tàu: kích thước thường rất nhỏ so với khung hình không ảnh, mật độ phân bố không đồng đều và thường xuyên bị nhiễu bởi sóng biển, mây che phủ hoặc các cấu trúc phức tạp ven bờ [3], [4].

Trong thập kỷ qua, các phương pháp phát hiện tàu đã có những bước tiến vượt bậc. Các nghiên cứu ban đầu dựa trên mô hình hai giai đoạn như R-CNN và các biến thể của nó đã đạt được độ chính xác đáng kể. Ví dụ, Zhang và cộng sự [5] đã đề xuất cải tiến Faster R-CNN kết hợp với SVM để xử lý vùng quan tâm, hay Zhang và cộng sự [6] phát triển OSCD-Net với cơ chế Deformable RoI để phát hiện hướng tàu. Tuy nhiên, các dòng mô hình này có cấu trúc phức tạp và tốc độ suy luận thấp, khiến chúng khó đáp ứng yêu cầu giám sát thời gian thực trên các thiết bị biên [7].

Gần đây, các kiến trúc một giai đoạn và Transformer đã thu hút sự chú ý nhờ khả năng cân bằng tốt hơn giữa tốc độ và độ chính xác. Công bố [8] giới thiệu FrequencyFormer sử dụng Transformer

trong miền tần số, trong khi công trình [9] đề xuất E-FPN dựa trên lý thuyết bằng chứng. Mặc dù cải thiện khả năng trích xuất đặc trưng toàn cục, các mô hình này thường yêu cầu tài nguyên tính toán (FLOPs) lớn. Để khắc phục, nhiều nghiên cứu tập trung vào cải tiến dòng mô hình YOLO. Nghiên cứu [10] đề xuất YOLO-LPSS tích hợp module CBAM và DySample để tăng cường phát hiện tàu nhỏ, [11] giới thiệu YOLO-StarLS sử dụng biến đổi Wavelet để khử nhiễu. Bên cạnh các kết quả đạt được, các phương pháp này thường mắc phải hai hạn chế lớn: (1) việc thêm các module phức tạp (Attention, Wavelet) làm tăng độ trễ tính toán và khó tối ưu hóa phần cứng; (2) Chúng vẫn phụ thuộc vào thuật toán Non-Maximum Suppression (NMS) ở khâu hậu xử lý. Trong bối cảnh giám sát cảng biển với mật độ tàu dày đặc, NMS trở thành "nút thắt cổ chai" gây độ trễ không ổn định và dễ bỏ sót các đối tượng nằm sát nhau [12]. Ngoài ra, việc sử dụng Distribution Focal Loss (DFL) trong các phiên bản như YOLOv8 hay YOLO11 cũng gây khó khăn cho quá trình lượng tử hóa và triển khai trên các hệ tính toán biên [7], [12].

Trong phạm vi nghiên cứu này, mục tiêu chính của nhóm tác giả là xây dựng một module phát hiện đối tượng tàu theo thời gian thực từ không ảnh, có thể tích hợp vào các hệ thống giám sát hàng hải tự động (ví dụ trên nền tảng UAV hoặc trạm xử lý ảnh vệ tinh), phục vụ các nhiệm vụ như giám sát xâm nhập trái phép, hỗ trợ tìm kiếm cứu nạn và quản lý hoạt động tàu thuyền ven bờ.

Để giải quyết triệt để các vấn đề trên, nghiên cứu này đề xuất ứng dụng mô hình YOLO26m [12], [13], [14], đây là kiến trúc tiên tiến thuộc thế hệ YOLO mới nhất cho bài toán phát hiện tàu từ không ảnh. Khác biệt với các tiếp cận trước đây vốn gia tăng độ phức tạp mô hình để đổi lấy độ chính xác [10], [11], YOLO26 sử dụng kiến trúc End-to-End NMS-Free, cho phép loại bỏ hoàn toàn độ trễ của khâu hậu xử lý và sự phụ thuộc vào các siêu tham số NMS [13]. Đồng thời, việc loại bỏ DFL kết hợp với các hàm mất mát tiên tiến như Soft-Target Anchor Loss (STAL) và Progressive Loss (ProgLoss) giúp mô hình tinh gọn hơn, dễ dàng triển khai trên phần cứng nhưng vẫn duy trì khả năng định vị chính xác các mục tiêu nhỏ [12].

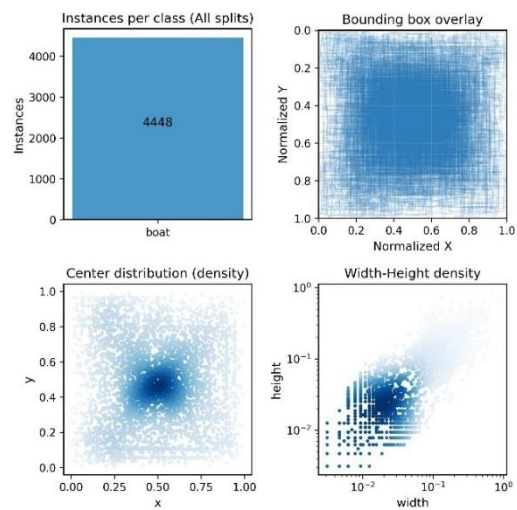
Nghiên cứu thực hiện huấn luyện chuyển tiếp và đánh giá phiên bản mô hình YOLO26m trên bộ dữ liệu "Kaggle Ships in Google Earth". Kết quả huấn luyện cho thấy phương pháp đề xuất đạt $mAP@0,5$ là 0,975, $mAP@0,5:0,95$ là 0,762 và tốc độ suy luận 14 ms/ảnh trên GPU Tesla P100, chứng minh hiệu năng ổn định so với các phương pháp hiện hành trong việc cân bằng

giữa độ chính xác và tốc độ xử lý thời gian thực.

2. Phương pháp thực hiện

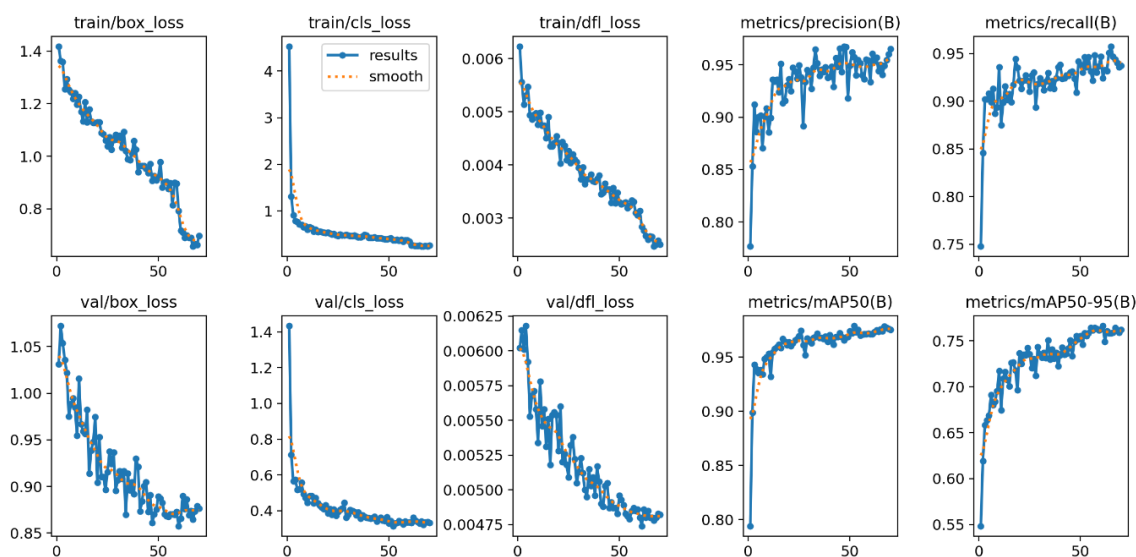
2.1. Bộ dữ liệu và cấu hình hệ thống

Bộ dữ liệu (dataset) được sử dụng là phiên bản "Kaggle Ships in Google Earth" được cung cấp thông qua Roboflow. Dataset gồm 1.293 ảnh vệ tinh quang học và ảnh chụp trên không ở các khu vực ven biển và vùng biển mở, trong đó có các loại tàu khác nhau. Các ảnh gốc có độ phân giải đa dạng và được cắt thành các ảnh vuông nhỏ hơn, đồng thời chuẩn hóa chuyển sang định dạng YOLO với tọa độ tương đối (x, y, w, h) .



Hình 1. Phân tích dataset

Kết quả phân tích dataset được thể hiện trong Hình 1, có 4,480 nhãn được đánh dấu với một lớp duy nhất là "boat", phù hợp với mô hình tập trung vào việc học đặc trưng hình học và vị trí của đối tượng tàu thuyền. Vị trí các tàu xuất hiện trên toàn bộ khung hình, nhưng có mật độ cao hơn quanh vùng trung tâm. Phân bố tọa độ tâm (x, y) tập trung gần $(0,5, 0,5)$ nhưng vẫn bao phủ các vùng biên, cho thấy sự đa dạng về vị trí và giúp mô hình học được đặc trưng cục bộ lẫn toàn cục. Phân bố kích thước tàu chủ yếu $w, h \in [0,01, 0,1]$ cho thấy đa số tàu có kích thước nhỏ so với ảnh, đồng thời tồn tại mối tương quan giữa chiều rộng và chiều cao, phản ánh hình dạng thon dài đặc trưng của tàu. Điều này là hợp lý với bài toán phát hiện đối tượng nhỏ, yêu cầu mô hình phải định vị chính xác và khai thác đặc trưng đa tỉ lệ. Dataset tuy có số lượng ảnh ở mức vừa phải (1.293 ảnh) nhưng với tổng số 4.480 nhãn "boat" và các kỹ thuật tăng cường dữ liệu được áp dụng sẽ giúp mô hình có đủ mẫu huấn luyện, có thể đạt được khả năng tổng quát hóa tốt trên tập kiểm chứng. Ngoài ra, với đặc điểm nhiều mục tiêu nhỏ, phân bố không



Hình 2. Diễn biến các chỉ số trong quá trình huấn luyện 70 epoch

gian đa dạng và hình dạng kéo dài, dataset này phù hợp cho việc huấn luyện bằng mô hình YOLO26m trong bài toán phát hiện tàu từ không ảnh.

Để phục vụ việc huấn luyện, dataset chia thành 03 tập nhỏ là Train, Test và Val theo tỉ lệ 70%/ 15%/ 15%. Điểm đáng chú ý, các ảnh trong một tập nhỏ có thể chứa nhiều tàu hoặc không chứa tàu nào, phản ánh đúng kịch bản giám sát vùng nước thực tế.

YOLO26m thuộc Ultralytics 8.4.7, tương thích Python 3.12.10, có 132 layer với 20.350.223 tham số và khoảng 67,8 GFLOPs. Mô hình này được sử dụng để phục vụ việc huấn luyện dataset trên môi trường Kaggle với hệ thống Linux x86_64, 4 lõi CPU, RAM 32GB và GPU Tesla P100-PCIE-16GB với Torch 2.x và CUDA hỗ trợ GPU Tesla P100. Các thông số này đảm bảo đủ không gian lưu trữ dataset, trọng số mô hình và thực thi việc huấn luyện.

2.2. Cấu hình huấn luyện chuyển tiếp YOLO26m

Phiên bản YOLO26m đã được tiền huấn luyện trên dữ liệu COCO sẽ được sử dụng cho huấn luyện chuyển tiếp. Các tham số chính như sau:

- Kích thước ảnh: 640×640.
- Số epoch: 70, dừng sau khi hội tụ.
- Batch size: 16.
- Optimizer: Áp dụng thuật toán SGD với momentum 0,937, hệ số regularization 0,0005, learning rate ban đầu 0,01.

Augmentation: Áp dụng các phương pháp tăng cường dữ liệu gồm mosaic, mixup, lật ảnh, dịch chuyển, thay đổi tỉ lệ và biến đổi không gian màu.

Mô hình được huấn luyện toàn phần, cho phép thích nghi cả backbone và head với đặc trưng đối tượng tàu thuyền trên nền mặt nước.

2.3. Đánh giá và kiểm định kết quả huấn luyện

Hiệu quả của mô hình trong quá trình huấn luyện được định lượng thông qua các chỉ số theo epoch gồm *box loss*, *classification loss*, *distribution focal loss (df_l loss)* trên tập huấn luyện và kiểm định, cùng với các chỉ số *precision*, *recall*, *mAP@0,5* và *mAP@0,5:0,95*. Các chỉ số này phản ánh quá trình tối ưu hoá và sự hội tụ của mô hình cũng như sự cải thiện về chất lượng dự đoán trong suốt 70 epoch.

Đồng thời, các đồ thị đánh giá như *Precision-Recall*, *Precision-Confidence*, *Recall-Confidence*, *F1-Confidence* và ma trận nhầm lẫn cho lớp “boat” được phân tích toàn diện khả năng phát hiện và phân loại của mô hình.

3. Kết quả quá trình huấn luyện

3.1. Sự hội tụ trong quá trình huấn luyện

Hiệu quả của quá trình huấn luyện được đánh giá thông qua sự biến thiên của các hàm mất mát và sự cải thiện của các chỉ số độ chính xác theo thời gian. Sự hội tụ của mô hình được thể hiện rõ nét qua biểu đồ các thành phần hàm mất mát trong Hình 2. Trên tập huấn luyện, giá trị *box loss* giảm đều từ 1,4 xuống 0,7, trong khi trên tập kiểm định, giá trị hội tụ quanh mức 0,87. Điều này cho thấy khả năng định vị tọa độ tàu của mô hình ngày càng chính xác. Giá trị *classification loss* giảm mạnh từ hơn 4,0 xuống dưới 0,8 trên tập huấn luyện và ổn định quanh mức 0,35 trên tập kiểm định.

Sự chênh lệch thấp giữa *train/loss* và *val/loss* chứng tỏ mô hình không gặp hiện tượng overfitting nghiêm trọng, xác nhận tính hiệu quả của các kỹ thuật tăng cường dữ liệu đã áp dụng. Giá trị *dfl loss* giảm từ 0,006 xuống 0,0025, phản ánh sự tinh chỉnh trong việc dự đoán phân phối xác suất của các cạnh *bouding box*.

Song song với việc giảm thiểu hàm mất mát, các chỉ số đánh giá chất lượng dự đoán cũng tăng trưởng ổn định theo số vòng huấn luyện. Trên tập kiểm định, sau quá trình 70 epoch, mô hình đạt $mAP@0,5 = 0,975$, $mAP@0,5:0,95 = 0,762$, cùng với $precision = 0,965$ và $recall=0,938$ cho lớp “boat”. Sự hội tụ nhanh chóng và ổn định của cả hàm *loss* lẫn các chỉ số *mAP* khẳng định sự phù hợp của kiến trúc YOLO26m và chiến lược huấn luyện chuyên tiếp đối với đặc thù dữ liệu tàu thuyền kích thước nhỏ và nền phức tạp trong các không ảnh.

3.2. Đánh giá định lượng trên tập kiểm định

Để đánh giá tính hiệu quả của mô hình YOLO26m, một quy trình đánh giá định lượng toàn diện đã được thực hiện dựa trên tập dữ liệu kiểm chứng độc lập gồm 159 hình ảnh chứa 357 đối tượng tàu (boat). Kết quả thực nghiệm được tổng hợp và phân tích dựa trên sự tương quan giữa các chỉ số đo lường chuẩn và các

đường cong đặc trưng về hiệu năng.

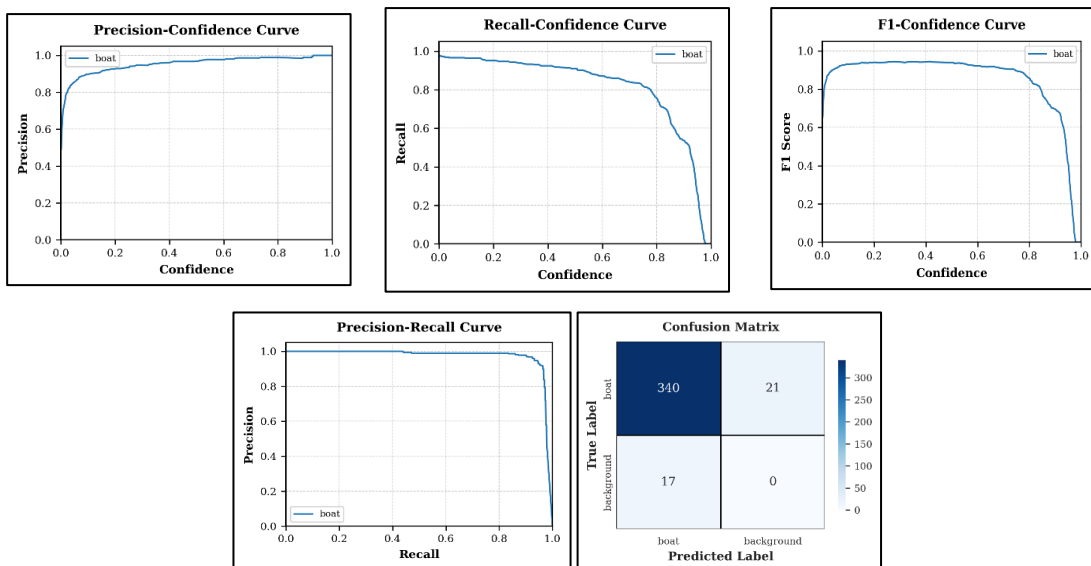
Các thông số chi tiết về hiệu năng định lượng kết quả kiểm tra được trình bày cụ thể trong Bảng 1. Mô hình đạt được các chỉ số đáp ứng yêu cầu xử lý thời gian thực với độ chính xác trung bình $mAP@0,5$ đạt 0,971 và $mAP@0,5:0,95$ đạt 0,766. Với số lượng tham số khoảng 20,35 triệu và khối lượng tính toán đạt 67,8 GFLOPs, mô hình YOLO26m cho thấy khả năng trích xuất đặc trưng đáp ứng tốt đối với không ảnh. Đặc biệt, hiệu năng thời gian thực được khẳng định qua tốc độ xử lý trung bình đạt 14,1 ms/ảnh, tương đương với tốc độ khung hình xấp xỉ 58,8FPS trên phần cứng Tesla P100.

Mối liên hệ giữa các ngưỡng tin cậy confidence và khả năng phân loại của mô hình được thể hiện chi tiết trong Hình 3. Đường cong *precision-confidence* và *recall-confidence* thể hiện xu hướng đối nghịch điển hình: khi tăng ngưỡng tin cậy, *precision* tăng dần và tiệm cận mức 1,0, trong khi *recall* có xu hướng giảm. Tuy nhiên, mô hình YOLO26m thể hiện tính ổn định cao khi đường cong Recall duy trì mức trên 0.8 cho đến ngưỡng tin cậy 0,75.

Sự giao thoa tối ưu giữa hai chỉ số này được xác định thông qua đường cong *F1-Confidence*, giá trị *f1-score*

Bảng 1. Tổng hợp các thông số hiệu năng định lượng của mô hình YOLO26m

Chỉ số đánh giá	Giá trị	Chỉ số tính toán	Giá trị
Precision	0.96	Tham số	20.35 M
Recall	0.924	GFLOPs	67.8
mAP@0.5	0.971	Suy luận	14.1 ms
mAP@0.5:0.95	0.766	Tốc độ khung hình	58.8



Hình 3. Hiệu năng mô hình YOLO26m trên tập kiểm định

đạt đỉnh tại 0,946 tương ứng với ngưỡng tin cậy 0,414. Tại điểm này, mô hình đạt được sự cân bằng tốt nhất giữa việc giảm thiểu sai lỗi nhận diện nhầm (dương tính giả) và bỏ sót đối tượng (âm tính giả). Việc đạt được giá trị *f1-score* cao trên một dải ngưỡng rộng (0,2 - 0,6) chứng minh mô hình có độ nhạy rất tốt và không quá phụ thuộc vào việc điều chỉnh ngưỡng khắt khe.

Khả năng phân biệt đối tượng của mô hình trên toàn bộ các ngưỡng hoạt động được minh chứng qua đường cong *Precision-Recall* (PR). Diện tích dưới đường cong PR đạt mức 0,971, phản ánh độ chính xác của mô hình trong việc tách biệt các đối tượng tàu ra khỏi các nhiễu nền phức tạp của môi trường biển. Đường cong này duy trì hình dạng tiệm cận góc trên bên phải, cho thấy ngay cả khi yêu cầu *recall* cao, độ *precision* của mô hình vẫn không bị suy giảm đáng kể.

Để làm rõ hơn về các sai số thực tế, *confusion matrix* cung cấp chi tiết hơn về kết quả phân loại trên tập 357 đối tượng *boat*. Mô hình đã nhận diện chính xác 340 trường hợp *boat*. Số lượng 21 *boat* bị bỏ sót chiếm tỷ lệ thấp, thường rơi vào các trường hợp tàu có kích thước siêu nhỏ hoặc bị che khuất bởi mây và sương mù. Đồng thời, chỉ có 17 trường hợp nhiễu nền bị phát hiện nhầm là *boat*, thể hiện khả năng lọc nhiễu ổn định của mô hình YOLO26m.

3.3. Đánh giá trực quan

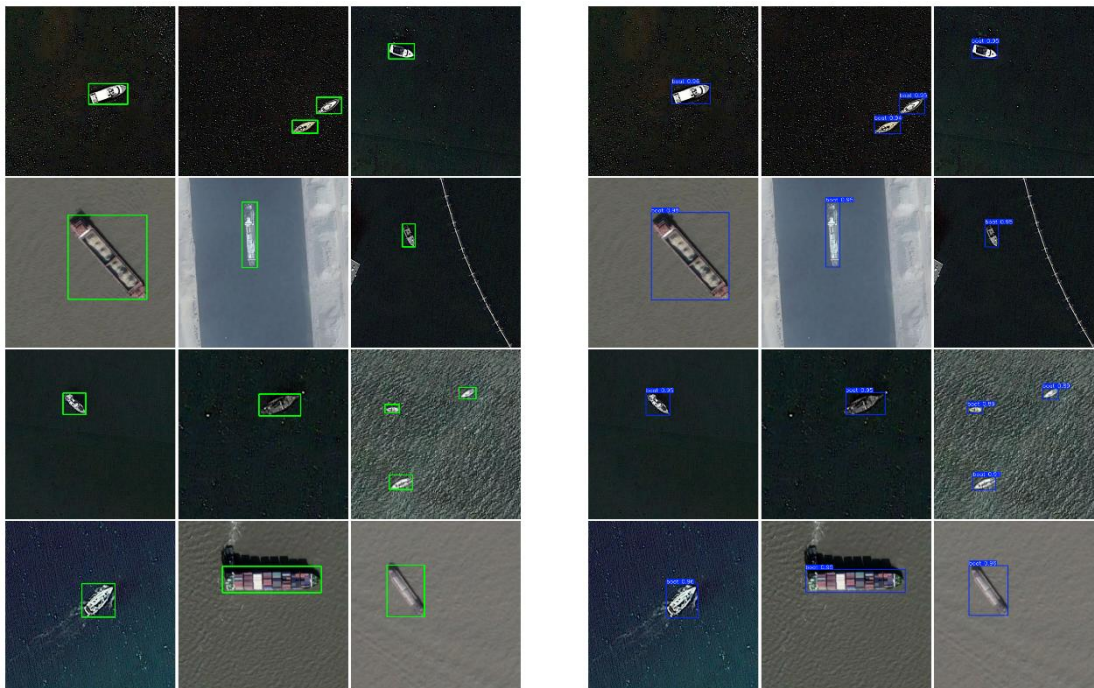
Hình 4(a) minh họa các mẫu ảnh từ bộ dữ liệu

kiểm định độc lập, với nhiều kịch bản thực tế như tàu đơn lẻ, mật độ cao, kích thước nhỏ và nền phức tạp (vùng ven bờ, công trình nhân tạo). Đây là cơ sở để đánh giá khả năng tổng quát hóa của mô hình. Hình 4(b) thể hiện kết quả thực tế mà mô hình YOLO26m đã phát hiện các đối tượng tàu đánh dấu bởi các *bounding box* màu xanh. YOLO26m cho thấy khả năng tách biệt đối tượng tốt, ngay cả khi các tàu nằm sát nhau hoặc có hướng quay đa dạng, không xảy ra hiện tượng gộp nhầm hay bỏ sót. Độ khớp cao của *bounding box* phù hợp chỉ số $mAP@0,5=0,971$.

Mặc dù vậy, ở các trường hợp tàu siêu nhỏ hoặc độ tương phản thấp, *bounding box* đôi khi chưa bám sát tuyệt đối biên tàu. Điều này lý giải sự khác biệt giữa $mAP@0,5$ và $mAP@0,5:0,95$, đồng thời chỉ ra tiềm năng cải thiện định vị ở các ngưỡng *IoU* cao trong các nghiên cứu tiếp theo. Tổng thể, kết quả trực quan khẳng định YOLO26m không chỉ mạnh về mặt thống kê mà còn vận hành ổn định trong môi trường thực tế phức tạp. Kết quả này củng cố tính khả thi khi ứng dụng mô hình vào các hệ thống giám sát hàng hải tự động dựa trên không ảnh.

4. Thảo luận

Hiệu năng của mô hình YOLO26m trong nghiên cứu này cho thấy sự xử lý tốt vấn đề cân bằng giữa độ chính xác và tốc độ xử lý đối với dữ liệu không ảnh. Chỉ số $mAP@0,5$ đạt 0,971 thể hiện mô hình có khả



a) Ảnh gốc

(b) Ảnh kết quả sau dự đoán

Hình 4. Kết quả thực nghiệm

năng tốt trong trích xuất đặc trưng hình thái tàu, ngay cả trong các điều kiện phức tạp như vùng ven bờ hoặc có cầu cảng.

Một điểm đáng chú ý là sự chênh lệch giữa $mAP@0,5$ (0,971) và $mAP@0,5:0,95$ (0,766). Điều này phản ánh việc duy trì độ khớp cao của *bounding box* ở các ngưỡng *IoU* khắt khe hơn vẫn gặp nhiều thách thức. Phân tích trực quan cho thấy nguyên nhân chủ yếu đến từ các đối tượng tàu có kích thước siêu nhỏ hoặc các tàu nằm sát cạnh nhau trong các bến cảng với mật độ cao, gây khó khăn cho việc định vị biên vật thể một cách tốt nhất.

Về dữ liệu huấn luyện, dataset “*Kaggle Ships in Google Earth*” chủ yếu phản ánh các kịch bản ven bờ, cảng và vùng biển mở với mật độ tàu và bối cảnh nhất định, do đó chưa thật sự bao phủ đầy đủ mọi loại phương tiện và kịch bản thực tế đặc thù như các hệ thống phao dày đặc hoặc đội hình tàu chuyên dụng. Tuy nhiên, các kết quả định lượng và trực quan đã trình bày cho thấy mô hình vẫn duy trì được hiệu năng tốt trong các bối cảnh phức tạp mà dataset hiện có tái hiện được.

Về mặt vận hành, $f1-score=0,946$ tại ngưỡng tin cậy 0,414 cho thấy mô hình có thể đạt hiệu năng tốt mà không đòi hỏi cấu hình ngưỡng quá cao. Với 20,35 triệu tham số, YOLO26m đáp ứng tốt trong xử lý các biến thể về hướng quay và kích thước của tàu mà không làm suy giảm tốc độ khung hình (duy trì ở mức 58,8FPS). Tốc độ này nhanh hơn đáng kể so với yêu cầu xử lý thời gian thực tiêu chuẩn (30FPS), tạo ra khả năng triển khai trên các hệ thống giám sát tự động tích hợp trên UAV hoặc các trạm thu phát dữ liệu vệ tinh có tài nguyên tính toán vừa phải.

5. Kết luận

Nghiên cứu đã thực hiện huấn luyện chuyển tiếp và đánh giá mô hình YOLO26m cho bài toán phát hiện tàu từ không ảnh. Các kết quả thực nghiệm định lượng khẳng định tính hiệu quả của mô hình với $precision=96%$, $mAP@0,5=0,971$. Những giá trị này minh chứng cho khả năng thích nghi tốt của mô hình với các đặc thù của không ảnh và ảnh viễn thám như góc nhìn thẳng đứng, đối tượng tàu kích thước đa dạng và nền nhiễu mật biển. Về mặt kỹ thuật, YOLO26m thể hiện sự tối ưu về chi phí tính toán khi chỉ tiêu tốn 67,8 GFLOP nhưng vẫn đảm bảo khả năng xử lý thời gian thực ở tốc độ cao. Kết quả trực quan cũng xác nhận mô hình có khả năng tổng quát hóa tốt trên nhiều kịch bản thực tế, từ các tàu đơn lẻ đến các khu vực bến cảng có mật độ tương đối dày đặc. Những kết quả này khẳng định YOLO26m là một lựa chọn phù hợp cho

các hệ thống giám sát hàng hải tự động, khi vừa đảm bảo độ chính xác cao vừa duy trì tốc độ suy luận nhanh.

Trong các nghiên cứu tiếp theo, nhóm tác giả dự kiến mở rộng đánh giá trên các dataset lớn hơn, hình ảnh thu sẽ từ nhiều nguồn đa dạng hơn (từ UAV hoặc ảnh vệ tinh) trong nhiều điều kiện môi trường khác nhau để kiểm chứng thêm khả năng tổng quát hóa của mô hình. Thứ hai, nhóm tác giả dự kiến thực hiện so sánh YOLO26m với các kiến trúc phổ biến khác như YOLOv5, YOLOv8 và Faster R-CNN trên cùng một dataset. Phép so sánh này sẽ giúp đánh giá toàn diện hơn sự đánh đổi giữa độ chính xác, tốc độ suy luận, số lượng tham số và chi phí tính toán, từ đó cung cấp cơ sở vững chắc cho việc lựa chọn kiến trúc phù hợp với các kịch bản triển khai khác nhau. Thứ ba, để tăng giá trị học thuật và ứng dụng thực tiễn, chúng tôi dự kiến mở rộng bài toán từ phát hiện một lớp “*boat*” thành phát hiện và phân loại loại tàu (tàu cá, tàu hàng, tàu khách, tàu tuần tra,...). Điều này sẽ giúp hệ thống không chỉ phát hiện mà còn nhận dạng loại phương tiện, phục vụ tốt hơn các nhiệm vụ giám sát và phân tích hàng hải.

TÀI LIỆU THAM KHẢO

- [1] M. Emara, Y. S. A. Abushawareb, and E. Schirli (2025), *Detection of Small and Medium Sized Ships in Satellite Images Using YOLO Models*, *Curr. Trends Comput.*, Vol.3, No.1, pp.17-27.
- [2] M. T. Nguyen, V. N. Tran, and X. T. Truong (2025), *Developing Object Detection Algorithm for Optoelectronic Systems on Surface Vessels Using Deep Learning Models*, *J. Sci. Tech.*, Vol.14, No.02, pp.89-103.
- [3] X. Hong, D. Fu, J. Tang, V. Lyne, M. Luo, and F. Su (2024), *Ship detection in reefs and deep-sea with medium-high resolution images*, *Geo-spatial Inf. Sci.*.
- [4] G. Liu, Y. Zhang, X. Zheng, X. Sun, K. Fu, and H. Wang (2014), *A New Method on Inshore Ship Detection in High-Resolution Satellite Images Using Shape and Context Information*, *IEEE Geosci. Remote Sens. Lett.*, Vol.11, No.3, pp.617-621.
- [5] S. Zhang, R. Wu, K. Xu, J. Wang, and W. Sun (2019), *R-CNN-Based Ship Detection from High Resolution Remote Sensing Imagery*, *Remote Sens.*, Vol.11, No.6, p. 631.
- [6] J. Zhang, R. Huang, Y. Li, and B. Pan (2022), *Oriented Ship Detection Based on Intersecting Circle and Deformable RoI in Remote Sensing Images*, *Remote Sens.*, Vol.14, No.19, p. 4749.

- [7] V. Sanikommu, S. P. MARRIPUDI, H. R. Yekkanti, R. Divi, R. Chandrakanth, and P. Mahindra (2025), *Edge computing for detection of ship and ship port from remote sensing images using YOLO*, Front. Artif. Intell., Vol.8, p. 1508664.
- [8] S. Liu, H. Wang, Z. Li, and P. Wei (2025), *FrequencyFormer: Oriented Object Detection with Frequency Transformer*, EAI Endorsed Trans. AI Robot., Vol.4.
- [9] Y. Dong, K. Xu, C. Zhu, E. Guan, and Y. Liu (2023), *E-FPN: Evidential Feature Pyramid Network for Ship Classification*, Remote Sens., Vol.15, No.15, p. 3916.
- [10] L. Shen, T. Gao, and Q. Yin (2025), *YOLO-LPSS: A Lightweight and Precise Detection Model for Small Sea Ships*, J. Mar. Sci. Eng., Vol.13, No.5, p. 925.
- [11] Y. Wang, S. Zhang, J. Xu, Z. Cheng, and G. Du (2025), *YOLO-StarLS: A Ship Detection Algorithm Based on Wavelet Transform and Multi-Scale Feature Extraction for Complex Environments*, Symmetry (Basel), Vol.17, No.7, p. 1116.
- [12] R. Sapkota and others (2025), *Ultralytics YOLO Evolution: An Overview of YOLO26, YOLO11, YOLOv8, and YOLOv5 Object Detectors for Computer Vision and Pattern Recognition*, arXiv Prepr. arXiv2510.09653.
- [13] R. Sapkota, R. H. Cheppally, A. Sharda, and M. Karkee (2026), *YOLO26: Key Architectural Enhancements and Performance Benchmarking for Real-Time Object Detection,*” arXiv Prepr. arXiv2509.25164, Vol.3.
doi: <https://doi.org/10.48550/arXiv.2509.25164>.
- [14] Ultralytics (2026), *YOLO26 vs. YOLOv10: The Evolution of End-to-End Object Detection*, Ultralytics YOLO Docs. Accessed: Jan. 30, 2026. [Online]. Available: <https://docs.ultralytics.com/compare/yolo26-vs-yolov10/>.

Ngày nhận bài:	04/02/2026
Ngày nhận bản sửa:	27/02/2026
Ngày duyệt đăng:	28/02/2026