

ĐÁNH GIÁ THỰC NGHIỆM HỆ THỐNG NHẬN DẠNG VÀ PHÂN LOẠI VẬT THỂ TRÊN MÔ HÌNH ROBOT DELTA ỨNG DỤNG YOLO11 VÀ HOMOGRAPHY

EXPERIMENTAL EVALUATION OF OBJECT RECOGNITION AND SORTING ON A DELTA ROBOT PLATFORM USING YOLO11 AND HOMOGRAPHY

**Đặng An Bình^{1,*}, Nguyễn Công Hùng¹, Nguyễn Ngọc Linh¹, Chu Đức Khoan¹,
Bùi Thanh Lâm², Lê Văn Dũng³**

¹Trường Cao đẳng nghề Kỹ thuật công nghệ

²Trường Cơ khí - Ô tô, Đại học Công nghiệp Hà Nội

³Trường Đại học Hùng Vương

*Email: danganbinhktcn@gmail.com

TÓM TẮT

Bài báo đánh giá thực nghiệm hệ thống nhận dạng và phân loại vật thể trên mô hình robot Delta ứng dụng YOLO11 và phép biến đổi homography. Hệ thống gồm camera RGB đặt cố định phía trên băng chuyền, máy tính xử lý ảnh chạy Python và YOLO11, PLC Siemens S7 1200, robot Delta ba bậc tự do và cơ cấu hút chân không. Tọa độ tâm vật thể thu được từ YOLO11 được chuyển sang tọa độ robot bằng homography rồi ghi vào vùng nhớ trao đổi của PLC thông qua cơ chế Valid/Ack. Đánh giá được tổ chức theo bốn mức tách biệt: nhận dạng đúng lớp, định vị trong ngưỡng $\pm 1,5$ mm, gắp thành công và thả đúng khay – cách làm cho phép xác định khâu nào gây lỗi, không chỉ ghi nhận kết quả cuối chu trình. Thử nghiệm 300 mẫu thuộc sáu lớp vật thể cho kết quả lần lượt là 96,3%; 95,3%; 94,0%; 93,0% và tỷ lệ chu trình hoàn chỉnh đạt 92,0%. Pipeline YOLO11 – homography – PLC đáp ứng yêu cầu thực hành gắp-phân loại trong môi trường đào tạo Cơ điện tử, với khâu cơ cấu hút là điểm yếu cần ưu tiên cải thiện.

Từ khóa: YOLO11; Homography; Robot Delta; Phân loại vật thể; PLC; Thị giác máy; Đánh giá thực nghiệm.

ABSTRACT

This paper experimentally evaluates an object recognition and sorting system on a Delta robot platform using YOLO11 and homography. The system consists of a fixed RGB camera above a conveyor, a Python/YOLO11 host computer, a Siemens S7-1200 PLC, a 3-DOF Delta robot, and a vacuum gripper. Object centroids detected by YOLO11 are mapped to robot coordinates via homography and written to a shared PLC data block using a Valid/Ack handshake. Evaluation is structured across four separate levels – correct class recognition, positioning within ± 1.5 mm, successful pick, and correct bin placement – so that the failure-contributing stage can be identified rather than simply reporting an end-to-end success rate. Tests on 300 samples across six object classes yielded 96.3%, 95.3%, 94.0%, 93.0%, and an overall cycle completion rate of 92.0%. The

YOLO11 – homography – PLC pipeline meets the requirements for pick-and-sort training exercises in mechatronics education; the vacuum gripper assembly is identified as the primary weak point requiring improvement.

Keywords: *YOLO11; Homography; Delta robot; Object sorting; PLC; Machine vision; Experimental evaluation.*

1. ĐẶT VẤN ĐỀ

Precision 0,956. Recall 1,000. mAP@0.5 bằng 0,995. Nhìn vào tờ kết quả của một mô hình nhận dạng vật thể, khó tìm điểm để phê bình – nhưng khi robot thật phải đưa giác hút xuống đúng tâm vật thể đang nằm trên băng chuyền chạy, những con số đó đột nhiên trở nên không đủ thuyết phục. Model accuracy đo hiệu năng ở cấp pixel. Tỷ lệ gấp thành công phụ thuộc vào sai số chuyển đổi tọa độ, độ trễ truyền thông, độ chính xác định vị robot và lực hút của cơ cấu – bốn thứ mà kết quả mAP không phản ánh.

Nền tảng của bài toán phát hiện đối tượng bằng mạng nơ-ron sâu được đặt ra trong tổng kết của LeCun và cộng sự [1]. Từ đó, hai hướng kiến trúc phân kỳ rõ: nhóm hai giai đoạn điển hình là Faster R CNN [2] tạo vùng đề xuất trước rồi phân loại, và nhóm một giai đoạn điển hình là SSD [3] và YOLO [4] dự đoán trực tiếp hộp bao từ ảnh đầu vào. Trong bài toán băng chuyền cần tốc độ suy luận cao, nhóm một giai đoạn chiếm ưu thế rõ ràng. YOLOv4 [5] đưa vào nhiều kỹ thuật tối ưu mà không hy sinh tốc độ đáng kể; YOLO11 do Ultralytics phát hành [6] tiếp nối hướng đó. Lý do chọn YOLO11 ở đây không phải vì nó đứng đầu bảng benchmark – mà vì bộ công cụ huấn luyện và triển khai của Ultralytics phù hợp với người dùng không chuyên sâu về học sâu, điều quan trọng trong môi trường đào tạo nghề.

Từ ảnh đến tọa độ robot là một bước thường bị xem nhẹ. Với vật thể nằm trên mặt

phẳng, tái dựng 3D đầy đủ là thừa; phép biến đổi homography đủ để ánh xạ tọa độ điểm từ hệ ảnh sang hệ robot mà không cần thêm cảm biến độ sâu [7]. OpenCV [8] cung cấp các hàm triển khai tương ứng. Sai số của phép biến đổi này – thường chỉ được ghi nhận là “chấp nhận được” – quyết định một phần không nhỏ tỷ lệ gấp thành công, nhất là với giác hút đường kính nhỏ và vật thể nhẹ có bề mặt không đều.

Nền tảng chấp hành là robot song song Delta do Clavel đề xuất năm 1988 [9], được phân tích chi tiết trong các công trình về robot song song của Merlet [10] và Tsai [11]. Với cấu trúc vòng kín ba nhánh, robot Delta cho tốc độ đáp ứng tốt và quán tính phần chuyển động nhỏ – đặc điểm phù hợp với bài toán gấp-đặt vật thể nhẹ. Đặc tính vùng làm việc và điều kiện không suy biến Jacobian [12] đã được kiểm tra trong thiết kế; bài báo này không lặp lại phần đó mà chỉ dùng robot như nền tảng chấp hành.

Phần lớn nghiên cứu về robot gấp-phân loại ứng dụng thị giác máy báo cáo kết quả ở một hoặc hai cấp độ – thường là accuracy nhận dạng hoặc tỷ lệ gấp cuối cùng. Cách trình bày đó không cho thấy khâu nào trong chuỗi gây lỗi nhiều nhất, và vì vậy không chỉ ra được hướng cải thiện cụ thể. Bài báo này đánh giá pipeline YOLO11–homography–PLC–robot Delta theo bốn mức tách biệt, từ nhận dạng đúng lớp đến thả đúng khay, với mục đích định vị rõ nguồn suy giảm hiệu năng.

Đóng góp 1: Quy trình nhận dạng và truyền tọa độ từ camera đến PLC qua vùng nhớ

trao đổi với cơ chế Valid/Ack, đảm bảo đồng bộ không đọc lặp vật thể.

Đóng góp 2: Đánh giá hệ thống theo bốn mức lỗi tách biệt – nhận dạng, định vị, gấp, thả – giúp định vị khâu yếu nhất trong chu trình phân loại.

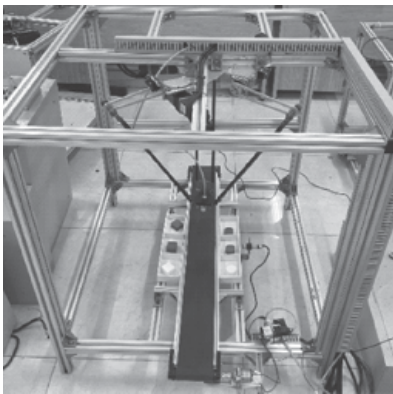
Đóng góp 3: Số liệu thực nghiệm trên 300 mẫu xác nhận pipeline YOLO11–homography–PLC phù hợp cho thực hành gấp–phân loại trong điều kiện đào tạo Cơ điện tử; đồng thời chỉ ra giới hạn cần khắc phục để nâng tỷ lệ chu trình hoàn chỉnh.

2. CẤU HÌNH HỆ THỐNG THỬ NGHIỆM

Hệ thống thử nghiệm gồm robot Delta ba bậc tự do, băng chuyền, camera RGB đặt cố định phía trên, máy tính xử lý ảnh, PLC Siemens S7 1200 [13], ba cụm servo Mitsubishi và driver [14], van khí nén và giác hút chân không. Camera hướng vuông góc với mặt phẳng băng, giúp giảm méo phối cảnh và đơn giản hóa bài toán homography. Cấu hình chi tiết ghi trong Bảng 1.

Bảng 1. Cấu hình hệ thống thử nghiệm.

Thành phần	Cấu hình sử dụng	Vai trò trong thử nghiệm
Robot	Delta 3-DOF	Gấp và di chuyển vật thể
Điều khiển	PLC Siemens S7-1200	Điều khiển logic và chấp hành
Truyền động	3 servo AC + hộp giảm tốc 1:25	Dẫn động ba khâu chủ động
Camera	RGB ≥ 2 MP, ≥ 30 FPS, gắn cố định	Thu ảnh vật thể trên băng chuyền
Xử lý ảnh	Python + YOLO11/YOLO11s	Nhận dạng lớp và tọa độ tâm
Giao tiếp	Ethernet / thư viện snap7	Truyền nhận, tọa độ, Valid/Ack
Cơ cấu gấp	Giác hút chân không Ø25 mm	Gấp vật thể nhẹ
Vật thể thử	6 lớp: tròn/vuông $\times$ đỏ/xanh/trắng	50 mẫu/lớp, tổng 300 mẫu



Hình 1. Cấu hình hệ thống thử nghiệm.

Sáu lớp vật thể được chọn theo nguyên tắc tạo các trường hợp dễ nhầm: cùng màu khác hình (tròn-đỏ với vuông-đỏ) và cùng hình khác màu (tròn-đỏ với tròn-xanh). Bố trí này giúp kiểm tra cả hai chiều phân loại của mô hình. Thử nghiệm được tổ chức theo ba cấp độ tách biệt – nhận dạng và homography trước (không chạy robot), định vị robot sau (không có vật thể), cuối cùng mới ghép chu trình đầy đủ – để tránh nhầm lẫn nguồn gây lỗi khi phân tích kết quả tích hợp.

3. PHƯƠNG PHÁP NHẬN DẠNG VÀ CHUYỂN ĐỔI TỌA ĐỘ

3.1. Nhận dạng vật thể bằng YOLO11

Ảnh đầu vào được chuẩn hóa về kích thước 640×640 trước khi đưa vào mạng. Đầu ra tại mỗi khung hình gồm danh sách hộp bao, mỗi hộp biểu diễn bởi bốn tọa độ góc:

$$B = (u_{\min}, v_{\min}, u_{\max}, v_{\max}) \quad (1)$$

Tọa độ tâm hộp bao – điểm được truyền sang hệ robot – xác định từ (1):

$$u_c = \frac{u_{\min} + u_{\max}}{2}, v_c = \frac{v_{\min} + v_{\max}}{2} \quad (2)$$

Một dự đoán chỉ được giữ lại khi điểm tin cậy vượt ngưỡng đặt trước:

$$p \geq p_{th} \quad (3)$$

Các hộp bao trùng lặp bị loại bởi Non-Max Suppression (NMS). Mức chồng lấn giữa hai hộp B_1 và B_2 đo bằng IoU:

$$IoU(B_1, B_2) = \frac{Area(B_1 \cap B_2)}{Area(B_1 \cup B_2)} \quad (4)$$

Hộp B_j bị loại khi:

$$B_j \text{ bị loại nếu } IoU(B_i, B_j) > \gamma \text{ và } p_i > p_j \quad (5)$$

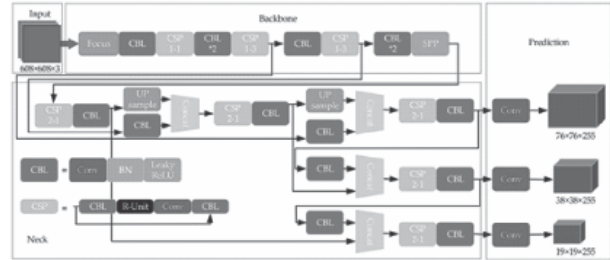
Hiệu năng mô hình được đo bằng Precision, Recall và F_1 :

$$Precision = \frac{TP}{TP+FP}, Recall = \frac{TP}{TP+FN}, F_1 = \frac{2 \cdot P \cdot R}{P+R} \quad (6)$$

Chỉ số mAP tính bằng diện tích dưới đường cong Precision-Recall trên tất cả các lớp:

$$mAP = \frac{1}{C} \sum_{c=1}^C \int_0^1 P_c(R) dR \quad (7)$$

mAP@0.5 tính với ngưỡng IoU bằng 0,5; mAP@0.5:0.95 là trung bình trên dải IoU từ 0,5 đến 0,95 [6]. Trong huấn luyện YOLO11, hàm mất mát gồm ba thành phần: hồi quy hộp bao: $\mathcal{L}_{box}$, phân loại: $\mathcal{L}_{cls}$ và phân phối biên: $\mathcal{L}_{dfl}$.



Hình 2. Pipeline phát hiện đối tượng bằng YOLO11.

3.2. Chuyển đổi tọa độ bằng homography

Với vật thể nằm trên mặt phẳng có chiều cao không đổi, homography đủ để ánh xạ tọa độ ảnh (u, v) sang tọa độ thực (X, Y) – không cần tái dựng 3D [7]:

$$s \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = H \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \quad (8)$$

Ma trận H (3×3) xác định từ tối thiểu bốn cặp điểm chuẩn trên bảng hiệu chỉnh, dùng hàm findHomography của OpenCV [8]. Sai số chuyển đổi tại điểm thứ i tính riêng theo từng trục:

$$RMSE_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \hat{X}_i)^2}, RMSE_y = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \quad (9)$$

RMSE tổng hợp:

$$RMSE_{xy} = \sqrt{\frac{1}{n} \sum_{i=1}^n [(X_i - \hat{X}_i)^2 + (Y_i - \hat{Y}_i)^2]} \quad (10)$$

Đánh giá riêng theo từng trục ở (9) giúp phát hiện sai số hệ thống theo một hướng nhất định – thường xảy ra khi camera bị nghiêng nhẹ hoặc băng chuyển không vuông góc chính xác với trục ảnh.

4. THUẬT TOÁN TRUYỀN DỮ LIỆU VÀ CHU TRÌNH PHÂN LOẠI

Sau khi nhận dạng và chuyển đổi tọa độ, Python ghi dữ liệu vào vùng nhớ DB của PLC theo nguyên tắc bắt tay có cờ đồng bộ. Chỉ

khi PLC đặt Ack = 1 Python mới ghi vật thể tiếp theo – cơ chế này ngăn PLC đọc lặp cùng một vật thể qua nhiều khung hình liên tiếp. Cấu trúc vùng nhớ và biến trạng thái ghi trong Bảng 2 và Bảng 3.

Bảng 2. Cấu trúc vùng nhớ trao đổi giữa Python và PLC.

Biến	Kiểu	Địa chỉ	Ý nghĩa
ObjectFlag	Byte	DB4.DBB0	Bit 0–5 tương ứng 6 lớp vật thể
X_pos	Real	DB4.DBD2	Tọa độ X sau homography (mm)
Y_pos	Real	DB4.DBD6	Tọa độ Y sau homography (mm)
Valid	Bool	DB4.DBX10.0	Python báo dữ liệu mới sẵn sàng
Ack	Bool	DB4.DBX10.1	PLC xác nhận đã đọc xong
ErrorCode	Word	DB4.DBW12	Mã lỗi truyền thông hoặc nhận dạng

Bảng 3. Biến trạng thái trong chương trình xử lý ảnh.

Biến	Kiểu	Ý nghĩa
label	int	Nhãn lớp vật thể (0–5)
confidence	float	Điểm tin cậy từ YOLO11 (0–1)
cx, cy	float	Tọa độ tâm hộp bao trên ảnh (pixel)
X, Y	float	Tọa độ thực sau homography (mm)
object_in_zone	bool	Vật thể nằm trong vùng gấp
Valid	bool	Cờ dữ liệu hợp lệ để ghi DB
Ack	bool	Cờ xác nhận từ PLC
frames_no_object	int	Số frame liên tiếp không phát hiện vật

Điều kiện để vật thể được chọn gấp – nằm trong vùng gấp quy ước:

$$|x_c - x_g| \leq \Delta x_g \quad (11)$$

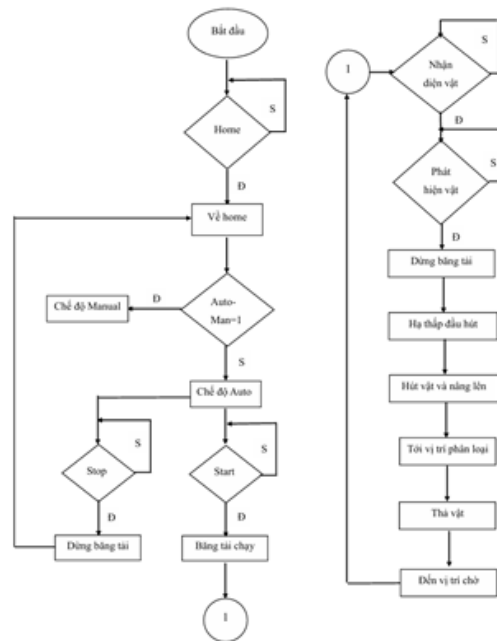
Trong đó, x_c là tọa độ tâm vật thể theo phương chuyển động băng, x_g là đường gấp và Δx_g là biên chấp nhận. Điều kiện (11) giảm sai số do vật thể tiếp tục di chuyển trong thời gian xử lý. Tổng trễ của chu trình từ khi thu ảnh đến khi PLC nhận tọa độ:

$$T_{\text{delay}} = T_{\text{cam}} + T_{\text{inf}} + T_{\text{post}} + T_{\text{comm}} + T_{\text{PLC}} \quad (12)$$

Sai lệch vị trí gấp do trễ:

$$\Delta x_{\text{delay}} = v_b \cdot T_{\text{delay}} \quad (13)$$

Quan hệ (13) có hàm ý thực tiễn: tốc độ băng chuyển v_b không thể chọn độc lập với tốc độ xử lý ảnh. Trong thử nghiệm, v_b giữ ở mức thấp đủ để Δx_{delay} không vượt bán kính giác hút.



Hình 3. Lưu đồ pipeline xử lý ảnh và truyền tín hiệu đến PLC.

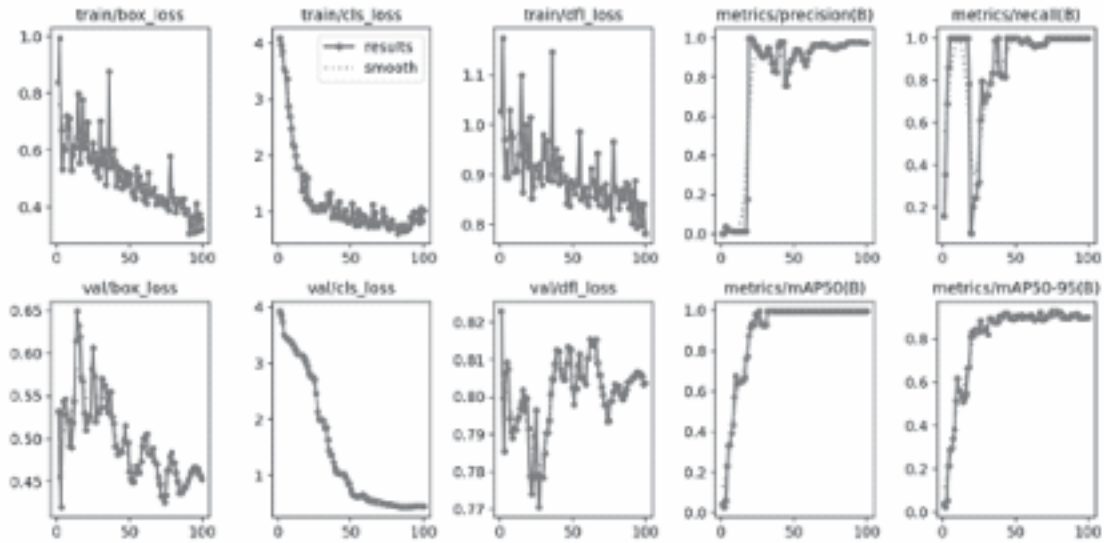
5. KẾT QUẢ NHẬN DẠNG VẬT THỂ

Mô hình YOLO11 được huấn luyện trên bộ dữ liệu ảnh chụp trong điều kiện phòng thực hành – cùng camera, cùng độ cao và nền băng chuyền thực tế. Tập kiểm thử hoàn toàn tách biệt với tập huấn luyện. Bảng 4 tổng hợp kết quả.

Bảng 4. Kết quả huấn luyện mô hình YOLO11 trên bộ dữ liệu 6 lớp.

Chỉ số	Giá trị	Ghi chú
Số lớp	6	Tròn/Vuông × Đỏ/Xanh/Trắng
Số epoch huấn luyện	100	Dừng khi validation loss hội tụ
Kích thước ảnh đầu vào	640×640	RGB, chuẩn hóa [0, 1]
Precision trung bình	0,956	Ngưỡng confidence 0,25; IoU 0,5
Recall trung bình	1,000	50 ảnh test/lớp × 6 lớp; xem chú thích *
F1-score trung bình	0,978	Tính từ Precision và Recall trung bình
mAP@0.5	0,995	IoU threshold 0,5
mAP@0.5:0.95	0,927	Trung bình dải IoU 0,50–0,95 bước 0,05
File trọng số	best.pt	Cần lưu kèm file cấu hình YAML

* Recall = 1,000 ghi nhận trên 300 ảnh kiểm thử ở ngưỡng confidence 0,25 và IoU 0,5. Nếu tăng ngưỡng IoU lên 0,75 hoặc giảm ngưỡng confidence, Recall sẽ giảm. Kết quả cần được hiểu trong phạm vi điều kiện kiểm thử cụ thể này.



Hình 4. Đường cong huấn luyện mô hình YOLO11.

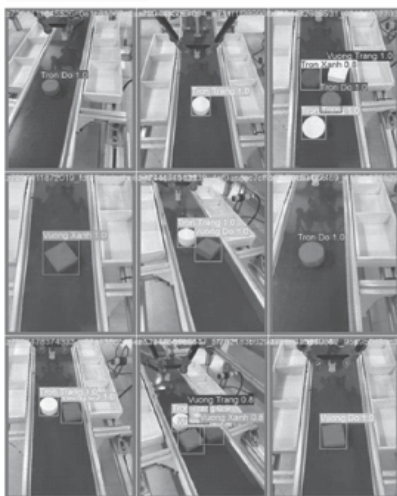
mAP@0.5 đạt 0,995 là mức cao với sáu lớp vật thể đơn giản – không đáng ngạc nhiên với bộ dữ liệu chụp ở điều kiện kiểm soát. Precision 0,956 thấp hơn Recall cho thấy mô hình có xu hướng tạo false positive hơn là bỏ sót, điều ảnh hưởng đến tỷ lệ truyền nhầm tọa độ sang PLC nếu không đặt ngưỡng confidence đủ cao. Hình 6 thể hiện ma trận nhầm lẫn; lớp Vương-Đỏ xuất hiện phần lớn nhầm lẫn do màu đỏ nhạy cảm với cường độ chiếu sáng.

6. KẾT QUẢ ĐỊNH VỊ VÀ CHUYỂN ĐỔI TỌA ĐỘ

Sai số định vị robot được đo tại năm điểm trong vùng làm việc trung tâm, mỗi điểm lặp 10 lần, theo tinh thần của ISO 9283 [15]. Sai số vị trí tại lần thử thứ k và giá trị trung bình:

$$e_{p,k} = \sqrt{(x_k - x_d)^2 + (y_k - y_d)^2 + (z_k - z_d)^2}, \bar{e}_p = \frac{1}{n} \sum_{k=1}^n e_{p,k} \quad (14)$$

Bảng 5 so sánh sai số chuyển đổi tọa độ giữa YOLO11s và phương pháp đường cơ sở dựa trên ngưỡng màu HSV kết hợp phát hiện đường viền (gọi tắt là Phương pháp HSV + Contour – không phải mô hình học sâu, chỉ dùng công cụ xử lý ảnh truyền thống của OpenCV). Cả hai phương pháp dùng cùng ma trận homography H, nên sai số RMSE phản ánh độ ổn định của tọa độ tâm được cung cấp, không phải độ chính xác của H.



Hình 5. Ảnh nhận dạng đúng và sai trên băng chuyền thực nghiệm.

Bảng 5. So sánh sai số chuyển đổi tọa độ và thời gian xử lý giữa YOLO11s và Phương pháp HSV+Contour.

Chỉ tiêu	YOLO11s (GPU)	HSV + Contour (CPU)	Ghi chú
Thời gian tiền xử lý	2,1 ms	3,4 ms	Resize + normalize vs resize only
Thời gian suy luận/phân ngưỡng	3,8 ms	8,2 ms	GPU inference vs CPU morphology
Hậu xử lý/NMS	0,9 ms	–	HSV + Contour không dùng NMS
Tổng trễ nhận dạng	6,8 ms	11,6 ms	Đủ cho băng chuyền tốc độ thấp
RMSE _x	0,63 mm	1,72 mm	Tâm hộp bao ổn định hơn tâm contour
RMSE _y	0,71 mm	1,84 mm	Cần đo lại khi thay camera
RMSE _{xy}	0,95 mm	2,51 mm	Ngưỡng gấp ±1,5 mm
Nhận dạng đúng, ánh sáng chuẩn	94,6%	78,4%	Tập kiểm thử độc lập
Nhận dạng đúng, ánh sáng ±30%	89,7%	52,1%	HSV suy giảm mạnh khi ánh sáng thay đổi

RMSE_{xy} của YOLO11s đạt 0,95 mm – dưới ngưỡng ±1,5 mm. Phương pháp HSV + Contour vượt ngưỡng này (2,51 mm), xác nhận tâm contour kém ổn định hơn tâm hộp bao của YOLO11 đặc biệt khi có biến động ánh sáng. Sai số định vị robot cộng với sai số chuyển đổi tọa độ giải thích tại sao tỷ lệ định vị trong ngưỡng ±1,5 mm ở chu trình tích hợp (95,3%) thấp hơn tỷ lệ nhận dạng đúng (96,3%).

Bảng 6. Kết quả kiểm tra sai số định vị robot Delta (50 lần đo tại 5 điểm).

Điểm thử	Tọa độ đặt (X; Y; Z), mm	Số lần lặp	Sai số TB (mm)	Sai số cực đại (mm)	Độ lặp lại (mm)
P1	0; 0; -390	10	0,62	1,10	0,28
P2	80; 0; -390	10	0,75	1,32	0,34
P3	-80; 0; -390	10	0,78	1,38	0,36
P4	0; 80; -400	10	0,84	1,46	0,39
P5	0; -80; -400	10	0,88	1,52	0,42
TB	–	50	0,77	1,52	0,36

7. ĐÁNH GIÁ CHU TRÌNH TÍCH HỢP

Thử nghiệm tích hợp thực hiện trên 300 mẫu, mỗi lớp 50 mẫu. Tỷ lệ phân loại đúng theo lớp ghi trong Bảng 7; đánh giá bốn mức và chu trình hoàn chỉnh ghi trong Bảng 8. Tỷ lệ chu trình hoàn chỉnh:

$$Acc_{cycle} = \frac{N_{cycle}}{N_{total}} \times 100\% = \frac{276}{300} \times 100\% = 92,0\% \quad (15)$$

Bảng 7. Kết quả phân loại 300 mẫu theo từng lớp vật thể.

Lớp vật thể	Số mẫu	Đúng	Sai	Tỷ lệ (%)
Tròn – Trắng	50	49	1	98,0
Tròn – Đỏ	50	47	3	94,0
Tròn – Xanh	50	49	1	98,0
Vuông – Trắng	50	50	0	100,0
Vuông – Đỏ	50	46	4	92,0
Vuông – Xanh	50	48	2	96,0
Tổng cộng	300	289	11	96,3

Bảng 8. Đánh giá bốn mức và chu trình hoàn chỉnh (300 mẫu).

Mức đánh giá	Số lần thử	Thành công	Tỷ lệ (%)	Nguyên nhân lỗi chính
Nhận dạng đúng lớp	300	289	96,3	Nhầm màu, vật ở biên ảnh
Định vị trong ngưỡng $\pm 1,5$ mm	300	286	95,3	Sai số homography, camera rung nhẹ
Gấp thành công	300	282	94,0	Lệch tâm giác hút, bề mặt vật không kín
Thả đúng khay	300	279	93,0	Mapping lớp-khay sai, tọa độ thả lệch
Chu trình hoàn chỉnh	300	276	92,0	Tổng hợp các mức trên

Bảng 9. Phân tích nguyên nhân lỗi theo từng khâu trong chu trình tích hợp.

Khâu lỗi	Số trường hợp	Tỷ lệ / 300 mẫu	Biểu hiện	Hướng khắc phục
Nhận dạng sai lớp	11	3,7%	Nhầm màu; vật ở biên ảnh bị cắt hộp	Bổ sung ảnh biên; ổn định chiếu sáng
Định vị ngoài ngưỡng	14	4,7%	Tọa độ lệch > 1,5 mm do camera dịch nhẹ	Gắn camera cứng hơn; kiểm tra điểm chuẩn trước mỗi ca
Gấp không thành công	18	6,0%	Giác hút lệch tâm hoặc bề mặt vật không kín	Căn lại offset tâm hút; kiểm tra vật mẫu
Thả sai khay	21	7,0%	Mapping lớp-bit-vị trí khay không đúng	Thêm bước xác nhận nhãn trước khi thả trong PLC
Chu trình thất bại	24	8,0%	Ít nhất một khâu trên thất bại	Xử lý theo thứ tự ưu tiên từ cơ cấu hút đến nhận dạng

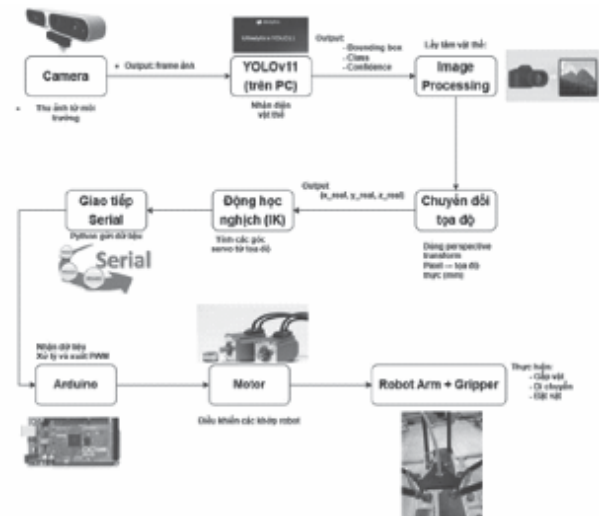
Ghi chú: Các nhóm lỗi không độc lập hoàn toàn – một mẫu có thể gặp lỗi ở nhiều khâu đồng thời. Số lỗi mỗi khâu tính độc lập; số chu trình thất bại là số mẫu có ít nhất một khâu thất bại.

Điều đáng chú ý ở Bảng 8 là khoảng cách giữa tỷ lệ nhận dạng (96,3%) và tỷ lệ chu trình hoàn chỉnh (92,0%). Chênh lệch 4,3 điểm phần trăm không đến từ model accuracy – nó đến từ chuỗi tích lũy lỗi nhỏ ở định vị, gấp và thả. Nếu chỉ cải thiện mô hình YOLO11, tỷ lệ chu trình sẽ không tăng đáng kể. Bảng 9 làm rõ điều này: khâu gấp (6,0% lỗi) và thả khay (7,0%) đóng góp nhiều hơn khâu nhận dạng (3,7%).

Lỗi nhầm lớp Vuông-Đỏ nhiều hơn các lớp khác. Ánh sáng không ổn định là nguyên nhân nhiều khả năng – màu đỏ biến thiên mạnh theo cường độ chiếu sáng trong không gian RGB. Với lớp này, giải pháp không phải là tăng số epoch huấn luyện, mà là bổ sung dữ liệu chụp ở nhiều điều kiện sáng khác nhau và có thể cân nhắc augmentation màu sắc khi huấn luyện lại.

Sai số homography tăng khi camera dịch chuyển nhẹ so với vị trí hiệu chỉnh ban đầu. Ba trong số 14 lỗi định vị ghi nhận ở các phiên vận hành sau khi camera bị chạm nhẹ. Kiểm tra lại điểm chuẩn trước mỗi phiên thử nghiệm dài – hoặc gắn camera bằng giá cứng hơn – sẽ loại bỏ nhóm lỗi này hiệu quả hơn bất kỳ điều chỉnh thuật toán nào.

Khâu gấp là điểm yếu chính của hệ thống hiện tại. Lệch tâm giác hút so với tọa độ tính từ homography, kết hợp với bề mặt vật không hoàn toàn phẳng, tạo ra lỗi hút không giữ được vật. Cách khắc phục trực tiếp là đo offset thực tế giữa tâm giác hút và tâm robot bằng phép đo cơ học, rồi bù offset đó vào tọa độ gấp trước khi truyền sang PLC.



Hình 6. Hệ thống robot Delta vận hành chu trình gấp và phân loại vật thể.

8. KẾT LUẬN

Bài báo đã đánh giá thực nghiệm pipeline YOLO11 – homography – PLC – robot Delta theo bốn mức tách biệt. Tỷ lệ nhận dạng đúng đạt 96,3%, RMSE chuyển đổi tọa độ đạt 0,95 mm (dưới ngưỡng gấp $\pm 1,5$ mm), và tỷ lệ chu trình hoàn chỉnh đạt 92,0% trên 300 mẫu sáu lớp vật thể. Chênh lệch 4,3 điểm giữa nhận dạng và chu trình hoàn chỉnh không đến từ mô hình YOLO11 mà đến từ chuỗi tích lũy lỗi ở định vị, hút và thả khay – điều chỉ có thể xác định bằng cách đánh giá từng mức tách biệt như trình bày ở bài này.

Phương pháp ngưỡng màu HSV + Contour không đáp ứng ngưỡng định vị $\pm 1,5$ mm với RMSE 2,51 mm, và suy giảm mạnh khi ánh sáng biến đổi. YOLO11s ổn định hơn ở cả hai điều kiện, khẳng định sự cần thiết của mô hình học sâu trong hệ thống nhận dạng theo thời gian thực khi ánh sáng không được kiểm soát chặt.

Hướng phát triển tiếp theo gồm: đo offset tâm giác hút và bù vào tọa độ gấp để giảm

lỗi cơ cấu hút; mở rộng dữ liệu huấn luyện cho lớp Vuông-Đô ở nhiều điều kiện chiếu sáng; thử nghiệm với tốc độ băng chuyền cao hơn kết hợp bù trễ theo (13); và tách đo từng thành phần trễ T_{inf} , T_{comm} riêng lẻ để xác định nút thắt trong công thức (12). Phương pháp đánh giá bốn mức trong bài này có thể được chuẩn hóa thành một bài thực hành đánh giá hệ thống tích hợp cho học phần Cơ điện tử. ❖

Lời cảm ơn:

Nhóm tác giả trân trọng cảm ơn Bộ Giáo dục và Đào tạo / Thiết kế, chế tạo mô hình robot song song ứng dụng trí tuệ nhân tạo phục vụ đào tạo nghề cơ điện tử. - CB2025-11 đã hỗ trợ kinh phí và tạo điều kiện thực hiện nghiên cứu này. ❖

Ngày nhận bài: 15/5/2026

Ngày phản biện: 26/5/2026

Tài liệu tham khảo:

- [1]. LeCun, Y., Bengio, Y., & Hinton, G. (2015), "Deep learning". Nature, 521(7553), 436-444. <https://doi.org/10.1038/nature14539>
- [2]. Ren, S., He, K., Girshick, R., & Sun, J. (2015), "Faster R-CNN: Towards real-time object detection with region proposal networks". Advances in Neural Information Processing Systems, 28, 91-99.
- [3]. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016), "SSD: Single Shot MultiBox Detector". Computer Vision – ECCV 2016, LNCS 9905, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2
- [4]. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016), "You only look once: Unified, real-time object detection". Proceedings of the IEEE CVPR 2016, 779-788. <https://doi.org/10.1109/CVPR.2016.91>
- [5]. Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020), "YOLOv4: Optimal speed and accuracy of object detection". arXiv:2004.10934. <https://doi.org/10.48550/arXiv.2004.10934>
- [6]. Ultralytics. (2024), Ultralytics YOLO11 Documentation. Ultralytics. <https://docs.ultralytics.com>
- [7]. Zhang, Z. (2000), "A flexible new technique for camera calibration". IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11), 1330-1334. <https://doi.org/10.1109/34.888718>
- [8]. Bradski, G. (2000), "The OpenCV library". Dr. Dobb's Journal of Software Tools, 25(11), 120-125.
- [9]. Clavel, R. (1988), "DELTA, a fast robot with parallel geometry". Proceedings of the 18th International Symposium on Industrial Robots, Lausanne, 91-100.
- [10]. Merlet, J.-P. (2006), "Parallel Robots" (2nd ed.). Springer. <https://doi.org/10.1007/1-4020-4133-0>
- [11]. Tsai, L.-W. (1999), "Robot Analysis: The Mechanics of Serial and Parallel Manipulators". John Wiley & Sons.
- [12]. Gosselin, C., & Angeles, J. (1990), "Singularity analysis of closed-loop kinematic chains". IEEE Transactions on Robotics and Automation, 6(3), 281-290. <https://doi.org/10.1109/70.56660>
- [13]. Siemens AG. (2025), SIMATIC S7-1200 Programmable Controller – System Manual (6ES7214-1AG40-0XB0). Siemens Industry Support.
- [14]. Mitsubishi Electric. (2013), HF-KN/HF-SN Servo Motor Instruction Manual. Mitsubishi Electric Factory Automation.
- [15]. ISO. (1998), ISO 9283:1998 – Manipulating industrial robots – Performance criteria and related test methods. International Organization for Standardization.