MỘT PHÂN TÍCH SO SÁNH CỦA YOLOV5, YOLOV8 VÀ

YOLOV10 AN TOÀN NHÀ BẾP

Athulya Sundaresan Geetha \* và Muhammad Hussain

Khoa Khoa học Máy tính, Đại học Huddersfield, Queensgate, Huddersfield HD1 3DH, Vương quốc Anh;

\*Thư từ: U2282847@unimail.hud.ac.uk;

Tháng Bảy 31, 2024

TRỪU TƯỢNG

An toàn dao trong nhà bếp là điều cần thiết để ngăn ngừa tai nạn hoặc thương tích với sự nhấn mạnh vào các phương pháp xử lý, bảo trì và lưu trữ thích hợp. Nghiên cứu này trình bày một phân tích so sánh của ba mô hình YOLO, YOLOv5, YOLOv8 và YOLOv10, để phát hiện các mối nguy hiểm liên quan đến việc xử lý dao, tập trung chủ yếu vào việc đảm bảo các ngón tay được cuộn tròn trong khi cầm các vật phẩm cần cắt và tay chỉ nên tiếp xúc với tay cầm dao tránh lưỡi dao. Độ chính xác, thu hồi, điểm F và ma trận nhầm lẫn chuẩn hóa được sử dụng để đánh giá hiệu suất của các mô hình. Kết quả chỉ ra rằng YOLOv5 hoạt động tốt hơn hai mô hình còn lại trong việc xác định mối nguy hiểm của việc đảm bảo tay chỉ chạm vào lưỡi dao, trong khi YOLOv8 xuất sắc trong việc phát hiện nguy cơ ngón tay cuộn tròn trong khi cầm đồ. YOLOv5 và YOLOv8 thực hiện gần như giống hệt nhau trong việc nhận dạng các lớp như tay, dao và rau, trong khi YOLOv5, YOLOv8 và YOLOv10 xác định chính xác thớt. Bài viết này cung cấp cái nhìn sâu sắc về những ưu điểm và thiếu sót của các mô hình này trong môi trường thực tế. Hơn nữa, bằng cách tối ưu hóa chi tiết các kiến trúc YOLO để xử lý dao an toàn, nghiên cứu này thúc đẩy sự phát triển của việc tăng độ chính xác và hiệu quả trong các hệ thống giám sát an toàn.

arXiv:2407.20872v1 [cs. CV] 30 Tháng Bảy 2024

*Từ khóa* Thị giác máy tính; YOLO; phát hiện đối tượng; Xử lý hình ảnh thời gian thực; Mạng nơ-ron tích chập; YOLOv5; YOLOv8; YOLOv10; An toàn dao

# Giới thiệu

Trong khi làm việc trong nhà bếp, một trong những tai nạn phổ biến nhất là tai nạn dao. Nó có thể xảy ra với ngay cả những đầu bếp có kinh nghiệm nếu con dao không được xử lý đúng cách với sự thận trọng tối đa. Phát hiện mối nguy hiểm khi làm việc với dao là rất quan trọng để ngăn ngừa tai nạn và đảm bảo tuân thủ các hướng dẫn an toàn. Xử lý dao an toàn bao gồm vị trí của bàn tay, ngón tay và dao; Trong bài báo này, trọng tâm chính là hai mối nguy hiểm bao gồm ngón tay cong và lưỡi chạm tay để tránh vết cắt.

Có nhiều thách thức khác nhau liên quan đến việc phát hiện các mối nguy hiểm trong việc xử lý dao an toàn. Sự thay đổi trong điều kiện ánh sáng trong nhà bếp có thể ảnh hưởng đến cách con dao xuất hiện trong hình ảnh. Nhà bếp thường lộn xộn với các thiết bị điện, đồ dùng và nguyên liệu, khiến con dao bị che khuất một phần hoặc hoàn toàn khỏi tầm nhìn. Đào tạo với một con dao duy nhất làm cho người mẫu quen với nó và không nhận ra các loại dao khác nhau. Chất lượng hình ảnh kém, liên quan đến hình ảnh có độ phân giải thấp hoặc mờ, khiến người mẫu khó tìm thấy dao. Chuyển động năng động của dao và tay có thể tạo ra những thách thức để phát hiện thời gian thực. Đôi khi, các dụng cụ trong nhà bếp có thể có hình dạng hoặc màu sắc tương tự như dao, có thể bị xác định sai là dao. Giải thích hiệu quả mối nguy hiểm, ví dụ, cầm dao sai, nhô ra các ngón tay được đặt trên vật phẩm cần thái hạt lựu hoặc gõ vào dao, là khá khó khăn đối với các mô hình.

Biểu đồ của gradient định hướng (HOG) được sử dụng để trích xuất các tính năng từ hình ảnh cả theo chiều ngang và chiều dọc, trong khi gradient định hướng biểu đồ dọc (VHOG) trích xuất các tính năng theo chiều dọc [1] và được phân loại theo máy vector hỗ trợ (SVM) và máy học cực đoan (ELM). Một kỹ thuật học máy, SVM, sử dụng trực quan hóa các hình ảnh và tính năng cục bộ trong các chuyển động động từ các khu vực không gian thời gian để phân loại nhưng không nhận ra các chuyển động [2, 3, 4].

Bởi vì các phương pháp tiếp cận hai giai đoạn mất rất nhiều thời gian, phụ thuộc vào các kiến trúc tiên tiến và cần giám sát thủ công thường xuyên, các mô hình CNN đã được giới thiệu để phát hiện và phân loại. R-CNN nhanh hơn cùng với GoogleNet hoạt động tốt hơn trong việc phát hiện các đối tượng [5]. Trong một nghiên cứu khác, ba thuật toán, MobileNet, MaskR-CNN và PoseNet, đã được sử dụng để cung cấp hiệu suất tốt hơn, phân đoạn hình ảnh và đánh giá mối nguy hiểm tương ứng. Để cải thiện nhận dạng đối tượng, nhiều mô hình như AlexNet [6], Google-Net [7], ResNet [8], VGG-Net [9], phương pháp Mạng tích chập dựa trên khu vực (R-CNN) [10], Fast-RCNN [11] và Faster-RCNN [12] đã được thiết kế [13].

Mặc dù hiệu suất được cải thiện, các vấn đề như xử lý dữ liệu hiệu quả, thực thi thời gian thực và kiến trúc tốt hơn vẫn tiếp tục. Đây là lúc các mô hình YOLO (You Only Look Once) trở nên hữu ích để khắc phục những nhược điểm mà quy trình phát hiện hai giai đoạn gặp phải; được xây dựng trên DarkNet, YOLOv1 và Fast YOLO lần lượt bao gồm 24 và 9 lớp tích chập [14]. YOLOv2 là phiên bản nâng cao, dựa trên DarkNet-19, với tiêu chuẩn hóa các lô, tham chiếu hộp, nhận dạng lớp độ sâu, các tính năng chi tiết và cụm kích thước cụ thể [15]. Sử dụng kiến trúc DarkNet-53, YOLOv3 kết hợp các mạng dư để trích xuất các tính năng, với entropy chéo nhị phân [16].

Cùng với đầu dự đoán của YOLOv3, YOLOv4 sử dụng CSPDarkNet53 làm kiến trúc xương sống và PANet với SPP làm mạng cổ để có hiệu suất tốt hơn [17].

Tốc độ hiệu suất của YOLOv5 đạt được bằng cách thêm một cổ cải tiến có tên CSP-PAN và một đầu có tên SPPF [18]. Ngoài tốc độ hiệu suất, YOLv6 làm giảm độ phức tạp của các yêu cầu tính toán [19], trong khi YOLOv7 với sự tích hợp của đầu phụ và đầu chì làm tăng độ chính xác trong việc phát hiện các đối tượng [20]. Trong YOLOv8, khung hỗ trợ nhận dạng đối tượng, phân đoạn hình ảnh và theo dõi chuyển động bằng xương sống CSPDarkNet53 và cổ PAN-FPN đã được sửa đổi. Giới thiệu Mạng tổng hợp lớp hiệu quả tổng quát và Thông tin chuyển màu có thể lập trình, cả hai mô hình, YOLOv9 và YOLOv10, đều cải thiện độ chính xác trong việc phát hiện các đối tượng; hơn nữa, trong YOLOv10, triệt tiêu không tối đa đã được loại bỏ [21, 22].

Mục đích của nghiên cứu này là tìm ra một phân tích so sánh về hiệu suất của YOLOv5, YOLOv8 và YOLOv10 trong việc xác định các mối nguy hiểm trong việc xử lý an toàn dao trong môi trường nhà bếp. Các chỉ số hiệu suất như độ chính xác, khả năng nhớ lại, mAP50, mAP50-95, điểm F1 và ma trận nhầm lẫn được đánh giá để tìm ra mô hình tốt nhất trong số ba. Điều này có thể hữu ích trong thời gian thực để cho người đó biết khi dao được sử dụng không đúng cách chủ yếu dựa trên các hướng dẫn an toàn để ngăn ngừa rủi ro.

Phần còn lại của bài viết như sau: Trong Tạp chí Văn học, các bài viết dựa trên xử lý dao đã được nghiên cứu và xây dựng đặc biệt trên cơ sở các mô hình YOLO. Quy trình làm việc và kiến trúc của các mô hình YOLO, YOLOv5, YOLOv8 và YOLOv10, được trình bày chi tiết trong phần có tiêu đề Phương pháp luận. Ma trận hiệu suất thăm dò được thảo luận trong phần Kết quả thử nghiệm và Thảo luận.

# Ôn tập văn học

Phát hiện các mối nguy hiểm để đảm bảo an toàn trong khi xử lý dao liên quan đến việc nhận ra các mối nguy hiểm tiềm ẩn và thực hiện các biện pháp để giảm thiểu rủi ro. Các máy vector hỗ trợ được sử dụng để phân loại ngoại tuyến sau khi thu thập các tính năng thị giác và chuyển động cục bộ xung quanh các vị trí không gian thời gian và chuyển động của chúng [2, 3]; Tuy nhiên, họ đã không nhận ra các hoạt động liên quan đến việc chuẩn bị thức ăn [4]. Do đó, bằng cách sử dụng các chỉ số cảm biến từ các thiết bị nhúng trong dao và thìa, một kỹ thuật đã được giới thiệu để xác định các hành động, cụ thể là thái hạt lựu, bóc vỏ, trộn và múc, bằng cách trích xuất thông tin về trung bình, năng lượng, phương sai và entropy, bao gồm cao độ và xoay để định vị thiết bị [23]. Trong khi bộ điều khiển Wii tích hợp được sử dụng trong tay cầm đồ dùng trong nhà bếp, một nghiên cứu khác không chỉ bổ sung bộ điều khiển Wii mà còn tích hợp các cảm biến này cùng với camera RGBD, được hướng về phía máy trạm, để cải thiện hiệu suất nhận dạng các hoạt động [24]. Hình ảnh được xử lý trước bằng cách sử dụng kích thước, hình dạng và vị trí liên kết trong hình ảnh để xác định máy dò dao bằng kỹ thuật thị giác máy tính, chẳng hạn như xác định đường viền, trích xuất cấu trúc và phân đoạn. Để phát hiện nhiều tính năng, bộ lọc Haar đã được sử dụng, trong đó dao được xác định chính xác trong 45% trường hợp và phân loại sai là dao trong 85% trường hợp [25]. Tuy nhiên, có những hạn chế như tiêu tốn nhiều thời gian hơn, kiến trúc cao cấp và giải thích liên tục của con người.

Các mạng thần kinh tích chập (CNN), chẳng hạn như AlexNet [26], Google-Net [7], ResNet [8], VGG-Net [9], phương pháp Mạng tích chập dựa trên khu vực (R-CNN) [10], Fast-RCNN [11] và Faster-RCNN [12] đã được đề xuất và có sự cải thiện đáng kể về thị giác máy tính để tăng cường phát hiện các đối tượng và phân loại hình ảnh [13]. Phát hiện dao và súng bằng cách sử dụng R-CNN nhanh hơn so sánh kiến trúc GoogleNet và SqueezeNet, cho thấy GoogleNet đạt được độ chính xác 46,68% để phát hiện dao và SqeezeNet hoạt động tốt hơn (85,44%) để phát hiện súng [5]. Faster R-CNN phát hiện đội mũ bảo hiểm bằng Retinex để có hình ảnh ngoài trời tốt hơn và K-means ++ cho mũ bảo hiểm nhỏ, thu được mAP là 94,3% và tốc độ phát hiện 11,62 hình ảnh mỗi giây [27]. Trong một nghiên cứu của Noever và Noever (2020) [28], để xác định dao và máy dò mối đe dọa, ba thuật toán đã được sử dụng, trong đó MobileNet phân loại với độ chính xác 95%, MaskRCNN phát hiện và phân đoạn tay khỏi dao, và PoseNet giúp định vị các điểm xương để cải thiện đánh giá mối đe dọa và làm rõ ý định.

2

Nghiên cứu đề xuất phương pháp phát hiện và phân loại an ninh hành lý bằng tia X với 6 hạng, theo đó YOLO đạt 97,4% đối với súng và ResNet đạt 73,2% đối với phát hiện dao[29]. Để phát hiện các đối tượng cho người khiếm thị trong môi trường mất trật tự, YOLOv2 đã được sử dụng để phát hiện và các phương pháp định hướng hình dạng đã được sử dụng để ước tính các mô hình ba chiều của các đối tượng trong nhà bếp và phòng cà phê, cung cấp các hướng dẫn an toàn. Với độ chính xác 90,45%, hệ thống xử lý ở tốc độ 0,86 khung hình / giây [30]. Bộ dữ liệu tương tác nhà bếp được sử dụng để phát hiện tay bằng YOLOv3, với VGG-16 đạt được Độ chính xác trung bình (AP) cao hơn là 62,2% khi so sánh với kiến trúc MobileNet-Lite; Những hạn chế là tay mờ và tiếng ồn [31]. Một mô hình YOLOv4 nhẹ cải tiến đã được đề xuất để phát hiện việc quản lý an toàn thời gian thực trong công việc điện tại chỗ. Để giảm độ phức tạp của mô hình và các tác vụ tính toán, nó đã sử dụng một mạng đường trục hiện đại với các kết cấu có thể tách rời theo chiều sâu và cấu trúc nút cổ chai đảo ngược di động; Để trích xuất nhiều tính năng và trích xuất các đối tượng nhỏ, một mạng hợp nhất hai chiều nâng cao đã được giới thiệu. Ngoài ra, độ chính xác của đào tạo và phát hiện đã được tăng cường bằng cách cải thiện sự mất tự tin. Nó đã giảm đáng kể số lượng thông số (93,11%), tăng tốc độ phát hiện (22%) và đạt được độ chính xác 84,82% trong việc tạo điều kiện an toàn trong thời gian thực [32].

Sử dụng YOLOv5 và Bộ lọc tương quan hạt nhân, SafeCOOK, một hệ thống thời gian thực, chi phí thấp, các thiết bị nhà bếp được theo dõi, tăng cường an toàn bằng cách xác định các tình huống nguy hiểm trong khi nấu. Chỉ riêng YOLOv5 đã mất các đối tượng chồng chéo; YOLOv5 với Theo dõi đối tượng đơn và Theo dõi nhiều đối tượng chỉ theo dõi một loại và nhiều đối tượng, tương ứng, nhưng có lỗi; và YOLOv5 với Theo dõi nhiều đối tượng và theo dõi bỏ học được xử lý nhất quán mà không có nhiều lỗi [33]. Để phát hiện chính xác xem nhân viên nhà bếp có đeo khẩu trang đúng cách hay không, mẫu YOLOv5 hoạt động tốt với mAP50 là 97,6%, với độ chính xác hoàn toàn trong 89 kỷ nguyên và tốc độ suy luận là 10 khung hình mỗi giây [34]. Các tác giả đã đề xuất một hỗ trợ thực tế tăng cường để nấu ăn bằng cách sử dụng các cảm biến tích hợp thông qua các chuyển động của bàn tay. Mặc dù mô hình YOLOv5 đạt được độ chính xác thấp hơn so với các mô hình, ResNet và ResNeXt, nhưng nó đã phân loại nhiều loại cùng một lúc. Các công việc trong tương lai sẽ tập trung vào việc cải tiến máy bằng cách giám sát nấu ăn theo thời gian thực, cải thiện việc xác định cử chỉ tay, phát triển các mô hình để phát hiện lỗi và tích hợp các mô hình để hỗ trợ phục hồi sau những sai lầm [35]. Phát hiện vật thể theo thời gian thực dựa trên YOLOv5 được phát triển để kiểm soát nguồn điện của lò sưởi hoặc cảnh báo người dùng bằng cách phát hiện các tình huống như theo dõi chất lỏng sôi, sắp xếp dụng cụ và bếp rỗng. YOLOv5 đạt được độ chính xác và khả năng thu hồi hơn 98%. Trong công việc trong tương lai, trọng tâm sẽ là sự tích hợp của vi điều khiển trong điện tử và mở rộng các loại đối tượng [36].

Hơn nữa, để đạt được tốc độ cao và độ chính xác tốt hơn trong việc xác định các vật phẩm liên quan đến lửa, các tác giả đã trình bày

Mô hình YOLOv6 để phát hiện cháy trong thành phố thông minh. YOLOv6 đã chứng minh hiệu suất cao (98%), thu hồi (96%) và độ chính xác (83%), xác định các vật thể liên quan đến lửa trong vòng 0,66 giây. Trong những năm tới, mục tiêu chính sẽ là cải thiện hiệu suất trong các kịch bản đầy thách thức và phát triển một mô hình, 3D CNN / U-Net, để sử dụng trong các lĩnh vực khác nhau [37]. Việc phát hiện vũ khí như dao và súng vì sự an toàn của con người được thực hiện bởi kiến trúc YOLOv7 và YOLOv8, cung cấp khả năng phát hiện chính xác vũ khí thông qua các chú thích hộp giới hạn. YOLOv7-e6 vượt trội hơn các mô hình khác bằng cách đạt được mAP cao 90,3% ở ngưỡng 0,5 IoU. Các tác phẩm trong tương lai sẽ khám phá YOLOv8 để tăng cường xác định các đối tượng trong thời gian thực trong các lĩnh vực khác nhau [38]. Một nghiên cứu đã sử dụng YOLOv8, lần đầu tiên thử nghiệm mô hình được đào tạo trước trên hình ảnh để phát hiện dao và đào tạo một mô hình tùy chỉnh để nhận dạng dao [39]. Trong một nghiên cứu khác, YOLOv8 đã được sử dụng để phát hiện các vi phạm vệ sinh, chẳng hạn như ăn mặc không phù hợp và sâu bệnh, với độ chính xác 89%, đảm bảo an toàn thực phẩm và sức khỏe của khách hàng [40].

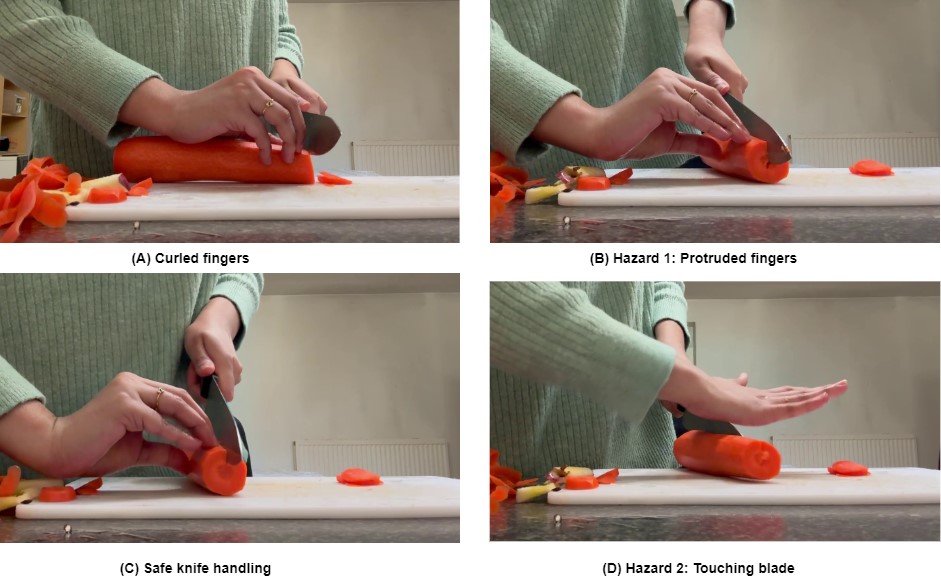
Để kết luận, tổng quan tài liệu này cung cấp ứng dụng các mô hình YOLO chủ yếu trong nhà bếp, từ phát hiện dao hiệu quả, cử động tay, vệ sinh, đến các biện pháp an toàn. Từ tất cả các mẫu YOLO, có một sự cải tiến không chỉ về hiệu suất và tốc độ mà còn về tính linh hoạt, ví dụ, hình ảnh bị mờ do chuyển động và các đối tượng nhỏ hoặc ẩn được xử lý tốt hơn với mỗi phiên bản mới. Mặc dù các kịch bản khác nhau đã được sử dụng trong môi trường nhà bếp, nhưng không có nghiên cứu nào về sự an toàn của việc xử lý dao. Bài viết này đề cập đến 2 mối nguy hiểm chính, ngón tay cong và lưỡi chạm tay, và so sánh hiệu suất của các mẫu YOLO, YOLOv5, YOLOv8 và YOLOv10, khi sử dụng dao.

# Phương pháp luận

## Tập dữ liệu

Bộ dữ liệu được lấy từ một video có độ phân giải cao (1920 × 1080 pixel) được quay bằng Apple iPhone15 Pro. Sau đó, video được xử lý và chuyển đổi thành 6004 khung hình riêng lẻ bằng tập lệnh Python. Các khung được trích xuất sau đó được chú thích thủ công bằng cách sử dụng Label Studio với 6 lớp sau: thớt, tay, rau, dao, nguy hiểm 1: cuộn ngón tay và nguy hiểm 2: lưỡi chạm tay. Việc sử dụng dao an toàn trong nhà bếp đã được kiểm tra, trong đó mối nguy hiểm 1 được chỉ định là ngón tay uốn để ngăn ngừa vết cắt vô tình và nguy cơ 2 được xác định là bàn tay tiếp xúc với lưỡi dao để tránh chấn thương tay nghiêm trọng. Hình ảnh mẫu của tập dữ liệu được thu thập được hiển thị trong Hình

1A-D, cho thấy vị trí ngón tay của bàn tay hỗ trợ, cuộn tròn (Hình 1A) và nhô ra (Hình 1B), cũng như xử lý dao an toàn (Hình 1C) và tiếp xúc tay với lưỡi dao (Hình 1D).



Hình 1: Hình ảnh mẫu của bộ dữ liệu nhà bếp.

## Tăng cường dữ liệu

Tăng cường dữ liệu hình ảnh là một kỹ thuật quan trọng trong giai đoạn tiền xử lý phân tích hình ảnh trong học máy và thị giác máy tính để cải thiện độ bền và hiệu suất của mô hình. Do kích thước hạn chế của tập dữ liệu, tăng cường hình ảnh đã được sử dụng để tạo ra nhiều hình ảnh bổ sung. Áp dụng các phép biến đổi đã giới thiệu sự thay đổi trong dữ liệu đào tạo, chuẩn bị các mô hình cho dữ liệu mới, không nhìn thấy. Với các hình ảnh đào tạo khác nhau, mô hình sẽ không rơi vào danh mục ghi nhớ các chi tiết cụ thể, do đó làm giảm tình trạng quá tải. Ngoài ra, tăng cường giải quyết sự mất cân bằng lớp bằng cách thêm nhiều hình ảnh hơn vào các lớp ít được trình bày, cải thiện độ chính xác của phân loại tổng thể và giảm sự thiên vị đối với các lớp đa số. Nhìn chung, tăng cường dữ liệu không chỉ làm tăng sự đa dạng của tập dữ liệu mà còn thúc đẩy sự phát triển của các mô hình học máy để giải quyết các khó khăn trong thế giới thực một cách hiệu quả.

Phát hiện các mối nguy hiểm trong nhà bếp trong khi xử lý dao đặt ra những thách thức cụ thể là điều kiện ánh sáng dao động, cử chỉ tay năng động, hình ảnh mờ, tầm nhìn bị cản trở và các góc độ đa dạng. Những thách thức này đã được giải quyết bằng cách kết hợp nhiều hình ảnh hơn vào bộ dữ liệu đào tạo. Bằng cách tăng cường hình ảnh với các biến thể trong điều kiện ánh sáng và tắc nghẽn tiềm năng, mô hình được đào tạo để thích ứng với các tình huống khác nhau. Điều này giúp mô hình hoạt động tốt hơn, tăng cường khả năng phát hiện các mối nguy hiểm. Ví dụ, các biến thể trong ánh sáng tạo ra các môi trường nhà bếp khác nhau, do đó chuẩn bị cho mô hình hoạt động trong các cài đặt ánh sáng khác nhau. Thêm tắc nghẽn bắt chước các thách thức trong thế giới thực, làm cho mô hình phát hiện các mối nguy hiểm ngay cả khi các vật thể bị cản trở một phần. Tăng cường đảm bảo rằng mô hình có khả năng xử lý sự phức tạp của môi trường nhà bếp, góp phần thực hành xử lý dao an toàn hơn.

### Cắt ngẫu nhiên

Cắt xén được sử dụng để cải thiện khả năng của mô hình để phát hiện các mối nguy hiểm trong khi xử lý dao an toàn, đảm bảo phát hiện khả năng hiển thị một phần của các đối tượng, chia tỷ lệ đồng đều hoặc khoảng cách khác nhau từ máy ảnh. Quá trình cắt xén điều chỉnh hình ảnh từ zoom tối thiểu 0%, giữ nguyên kích thước ban đầu, đến zoom tối đa 20%, tỷ lệ phần trăm kích thước hình ảnh để cắt từ mỗi bên. Do đó, kích thước của hình ảnh dao động từ 512 × 512 đến 640 × 640 pixel, điều chỉnh các biến thể trong thế giới thực trong việc sử dụng dao. Phương trình 1 được sử dụng để cắt 20% từ mỗi bên của hình ảnh.

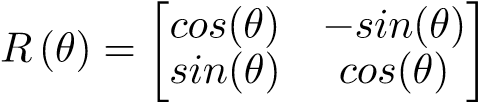
*D*′ = *D*[*đỉnh* + *p* × *chiều cao, đáy* − *p* × *chiều cao, bên trái* + *p* × *chiều rộng, bên phải* − *p* (1)

*D*′ là hình ảnh kết quả; *D* là hình ảnh gốc; và *p* đại diện cho tỷ lệ cắt xén là 20%.

### Xoay ngẫu nhiên

Xoay trong tập dữ liệu này liên quan đến việc thay đổi vị trí của dao trong một phạm vi xác định (tức là ±15 độ) (Phương trình 2). Kỹ thuật này giúp tăng cường độ chắc chắn của mô hình để phát hiện dao từ các góc độ khác nhau, tăng cường khả năng phát hiện và xử lý các mối nguy hiểm liên quan đến dao một cách hiệu quả trong các tình huống nhà bếp.

|  |  |
| --- | --- |
| *D*′ = *xoay*(*D,θ*)  Phương trình 3 hiển thị chuyển đổi quay 2 chiều của hình ảnh. | (2) |

 (3)

Ma trận quay được biểu diễn bằng *R*(*θ*).

### Cắt ngẫu nhiên

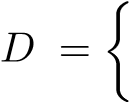
Kỹ thuật này liên quan đến việc nghiêng hình ảnh dọc theo trục x hoặc y, tạo ra độ nghiêng làm biến dạng các đối tượng trong ảnh. Để sử dụng dao an toàn, cắt ngẫu nhiên cung cấp các điều kiện trong đó dao hoặc tay xuất hiện nghiêng (±10 độ) do các góc khác nhau, dẫn đến biến dạng hình học (Phương trình 4).

*D*′ = *D* · *S* (4)

Ma trận cắt được ký hiệu là *S*.

### Thang độ xám ngẫu nhiên

Tỷ lệ xám, được áp dụng cho 15% hình ảnh màu, cải thiện hiệu suất của mô hình bằng cách xem xét các hình dạng và kết cấu thay vì màu sắc. Kỹ thuật này được sử dụng để cung cấp các điều kiện ánh sáng khác nhau, hình ảnh ánh sáng yếu hoặc đơn sắc, để cải thiện khả năng phát hiện các mối nguy hiểm của mô hình trong các cài đặt khác nhau (Phương trình 5).

 ′ 0*.*299 × *R* + 0*.*587 × *G* + 0*.*114 × *B,* với xác suất *p*

(5)

*D,* với xác suất (1 − *p*)

*R* biểu thị màu đỏ, *G* có nghĩa là màu xanh lá cây và *B* biểu thị màu xanh lam; *p* đại diện cho xác suất biến đổi.

### Huế

Huế đã được sử dụng để có được các biến thể màu sắc khác nhau trong hình ảnh. Phạm vi từ -15% đến + 15% biểu thị các biến thể về màu sắc kéo dài từ giá trị cơ sở -15% đến giá trị tối đa là 15% (Phương trình 6).

*D*′ = *D* + *H* (6)

Trong công thức tăng này, *H* là hệ số dịch chuyển màu sắc dao động từ -15% đến + 15%.

### Bão hòa

Độ bão hòa từ -25% đến + 25% phơi bày hình ảnh với các cường độ màu khác nhau, cung cấp các cài đặt ánh sáng và biến thể màu khác nhau (Phương trình 7).

*D*′ = adjust\_saturation(*D,α*) (7)

*α* là hệ số điều chỉnh độ bão hòa.

### Độ sáng

Ở đây, độ sáng đã được điều chỉnh trong phạm vi từ -15% đến +15% để phơi sáng hình ảnh với các điều kiện ánh sáng khác nhau (Phương trình 8). Điều này giúp mô hình thích ứng với các mức độ sáng khác nhau để phát hiện chính xác các mối nguy hiểm trong các môi trường khác nhau.

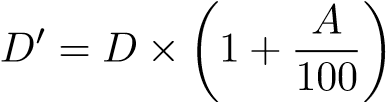
*D*′ = *cắt*(*D* + *β* × 255*,* 0*,*255) (8)

*β* là một yếu tố điều chỉnh độ sáng.

### Tiếp xúc

Phơi sáng có nghĩa là tăng hoặc giảm lượng ánh sáng thu được trong hình ảnh. Bằng cách điều chỉnh độ phơi sáng từ

-10% đến + 10%, nó làm giảm độ sáng lên đến -10% và tăng độ sáng lên + 10% so với hình ảnh gốc (Phương trình 9).

 *,* (9)

trong đó *A* là hệ số điều chỉnh phơi sáng (-10% đến +10%).

### Mờ

Làm mờ áp dụng hiệu ứng làm mờ cho hình ảnh để nâng cao hiệu suất của người mẫu nhằm tạo ra các cài đặt trong thế giới thực trong đó hình ảnh có thể không phải lúc nào cũng sắc nét và tương phản. Vì vậy, mô hình học cách phát hiện các mối nguy hiểm ngay cả trong các vật thể mất nét, dao hoặc tay, để xác định rủi ro trong các tình huống khác nhau. Nó cũng làm giảm tiếng ồn cũng như tăng số lượng các tính năng được trích xuất. Độ lệch chuẩn của hạt nhân mờ Gaussian được minh họa bằng *σ* (Phương trình 10).

*D*′ = *Mờ*(*D,σ*)*,* (10)

### Tiếng ồn ngẫu nhiên

Thêm nhiễu vào hình ảnh giúp mô hình tìm hiểu các tính năng quan trọng ngay cả khi có biến dạng (Phương trình 11). Theo phương pháp này, 0, 1% pixel trong hình ảnh bị ảnh hưởng bởi nhiễu.

*D*′ = *D* + *N* (11)

### HSV

HSV là viết tắt của hue, saturation và value, trong đó màu sắc biểu thị loại màu, độ bão hòa biểu thị cường độ màu và giá trị có nghĩa là độ sáng của hình ảnh. Giá trị được đặt thành 0,4 để giảm độ sáng xuống 40%. Hue được đặt thành 0,015 để tạo ra các biến thể màu sắc. Độ bão hòa đặt ở mức 0,7 để tăng độ bão hòa lên 70% để tăng cường độ màu. *σ* là độ sáng được điều chỉnh trên pixel (Phương trình 12).

*D*′ = *αD* (12)

### Dịch

Dịch dịch chuyển hình ảnh dọc theo trục dọc và trục ngang. Với dịch được đặt thành 0,1, hình ảnh có thể được định vị lại ở chiều rộng và chiều cao 10%, theo chiều ngang và chiều dọc. Các yếu tố dịch thuật được biểu diễn bằng *tx* và *ty* (Phương trình 13).

*D*′ (*x* + *Tx* × *chiều rộng, y* + *ty* × *chiều cao*) = *D*(*x, y*) (13)

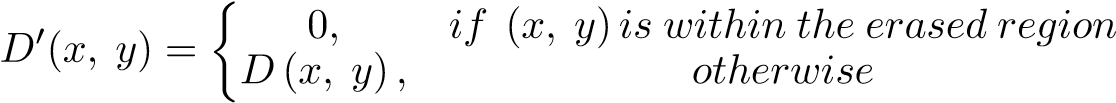
### Mosaic

Tăng cường khảm kết hợp bốn hình ảnh khác nhau thành một hình ảnh khảm duy nhất. Khi khảm = 1.0 được đặt, điều đó có nghĩa là áp dụng khảm cho 100% hình ảnh đào tạo để tăng cường sự đa dạng và phức tạp của dữ liệu đào tạo. Tọa độ hình ảnh khảm được biểu thị bằng (*x*′) và (*y*′); tọa độ tương ứng của hình ảnh gốc được ký hiệu là *xi* và *yi*; cường độ của hình ảnh được đặt là *Ii* (Phương trình 14).

*D*′ (*x*′*, y*′) = *Di*(*xi, ytôi*) (14)

### Xóa ngẫu nhiên

Bằng cách sử dụng phương pháp xóa, các phần của hình ảnh có thể bị xóa hoặc ẩn. Ở đây, giá trị 0, 4 biểu thị rằng 40% hình ảnh bị xóa. Sau quá trình xóa, cường độ của pixel được biểu thị bằng *D*′(*x, y*) (Phương trình 15).

 (15)

Tăng cường dữ liệu là rất quan trọng trong việc phát hiện các mối nguy hiểm khi sử dụng dao, vì nó giúp tạo ra các bộ dữ liệu lớn và tránh quá mức bằng cách tạo các kịch bản khác nhau để mô hình học hỏi.

## Kiến trúc YOLOv5

Kiến trúc của mô hình YOLOv5 bao gồm xương sống, cổ và đầu. Trong xương sống, YOLOv5 sử dụng mạng lưới một phần giai đoạn chéo (CSP) dựa trên ResNet, tăng hiệu quả thông qua các kết nối từng phần chéo giai đoạn và thực hiện nhiều khối tổng hợp kim tự tháp không gian (SPP) để trích xuất nhiều tính năng, giảm tải các tác vụ tính toán. Một mô-đun mạng tổng hợp đường dẫn (PAN) và các lớp nâng cấp được tích hợp vào cổ để tăng độ phân giải của bản đồ tính năng và tối ưu hóa việc trao đổi luồng thông tin giữa các lớp. Bao gồm 3 lớp chập, người đứng đầu YOLOv5 được sử dụng để dự đoán các hộp giới hạn, dự đoán lớp và điểm tin cậy [41].

Dự đoán dựa trên neo được sử dụng để liên kết hộp giới hạn với các hộp neo được xác định trước với các kích thước và biểu mẫu cụ thể. Hàm tổn thất trong YOLOv5 có Nhị phân chéo Entropy cho tổn thất lớp và đối tượng, và Giao điểm hoàn chỉnh trên Liên minh (CIoU) cho tổn thất vị trí, với phương pháp tính toán tổn thất đối tượng phụ thuộc vào kích thước của lớp dự đoán [13, 41]. Kiến trúc YOLOv5 bao gồm các lớp với một số bộ lọc, kích thước chỉ định kích thước bộ lọc và lặp lại hiển thị số lần lớp được sử dụng được đưa ra trong Bảng 1.

Bảng 1: Kiến trúc YOLOv5.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Lớp | Kích hoạt | Bộ lọc | Kích thước | Lặp lại | Kích thước đầu ra |
| Ảnh | - | - | - | - | 640 × 640 |
| Conv0 | ReLU | 16 | 3 × 3/2 | 1 | 320 × 320 |
| Conv1 | ReLU | 32 | 3 × 3/2 | 1 | 160 × 160 |
| Conv2 | ReLU | 64 | 3 × 3/2 | 1 | 80 × 80 |
| Conv3 | ReLU | 128 | 3 × 3/2 | 1 | 40 × 40 |
| Conv4 | ReLU | 256 | 3 × 3/2 | 1 | 20 × 20 |
| C3 | ReLU | 128 | 1 × 1 | 1 | 20 × 20 |
| Conv5 | ReLU | 128 | 3 × 3 | 1 | 20 × 20 |
| C3 | ReLU | 256 | 1 × 1 | 1 | 20 × 20 |
| Conv6 | ReLU | 256 | 3 × 3 | 1 | 20 × 20 |
| SPP | - | - | - | - | 20 × 20 |
| C3 | ReLU | 512 | 1 × 1 | 1 | 20 × 20 |
| Upsample | - | - | - | 1 | 40 × 40 |
| C3 | ReLU | 256 | 1 × 1 | 1 | 40 × 40 |
| Upsample | - | - | - | 1 | 80 × 80 |
| C3 | ReLU | 128 | 1 × 1 | 1 | 80 × 80 |
| Conv7 | ReLU | 128 | 3 × 3 | 1 | 80 × 80 |
| C3 | ReLU | 64 | 1 × 1 | 1 | 80 × 80 |
| Conv8 | ReLU | 64 | 3 × 3 | 1 | 80 × 80 |
| C3 | ReLU | 32 | 1 × 1 | 1 | 80 × 80 |

Cụ thể, mô hình sử dụng một loạt các lớp tích chập, với kích hoạt ReLu để giảm kích thước từ 640 × 640 xuống 320 × 320 xuống 20 × 20 pixel. Sau đó, lớp C3 bao gồm sự kết hợp của 1 × 1 và 3 × 3 kết cấu lặp đi lặp lại nhiều lần, mà không thay đổi kích thước của đầu ra. Tập hợp lớp kim tự tháp không gian (SPP) tập hợp các tính năng từ các quy mô khác nhau trong khi vẫn duy trì kích thước từ 20 × 20 pixel. Nâng cấp làm tăng ánh xạ tính năng lên độ phân giải cao hơn.

Khi so sánh với các phiên bản trước, YOLOv5 nhằm mục đích tăng hiệu suất và độ chính xác, giới thiệu cải tiến trong việc trích xuất nhiều tính năng, tổng hợp và dự đoán dựa trên các neo. Để làm cho nó thích ứng với các thiết bị iOS, các khung ONNX và CoreML đã được sử dụng thay vì PyTorch. Đạt được độ chính xác trung bình 50,7% trên bộ dữ liệu MS COCO, YOLOv5x đã xử lý hình ảnh 640 pixel ở tốc độ 200 khung hình / giây, với độ chính xác trung bình cao hơn là 55,8% ở kích thước hình ảnh 1534 pixel.

## Kiến trúc YOLOv8

Dựa trên YOLOv5, YOLOv8 kết hợp các cải tiến kiến trúc để có độ chính xác phát hiện vượt trội. YOLOv8 sử dụng mô-đun C2f để chuyển sang kiến trúc không neo, giúp đơn giản hóa việc dự đoán các hộp giới hạn và giảm sự phụ thuộc vào triệt tiêu không tối đa bằng cách loại bỏ các hộp neo truyền thống trình bày sai các mẫu khác nhau trong tập dữ liệu [41]. Bảng kiến trúc không chỉ cung cấp chi tiết cấu hình của từng lớp, kết hợp mô-đun C2f và Phân số gộp kim tự tháp không gian (SPPF) để tăng độ chính xác của vị trí và biểu diễn tính năng, mà còn nhấn mạnh loại và kích thước của bộ lọc và kích thước của đầu ra.

Mô hình xử lý hình ảnh đầu vào thông qua một chuỗi các lớp tích chập, Conv0, Conv2, C2f2 và C2f4, sử dụng kích hoạt ReLU để giảm kích thước pixel từ 320 × 320 xuống 80 × 80. Hơn nữa, kích thước được giảm xuống còn 20 × 20 trong Conv5 thành Conv7. Các lớp nâng cấp tăng kích thước từ 20 × 20 lên 80 × 80 pixel. Cuối cùng, các lớp tích chập còn lại giúp phát hiện đối tượng (Bảng 2).

Bảng 2: Kiến trúc YOLOv8.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Lớp | Kích hoạt | Bộ lọc | Kích thước | Lặp lại | Kích thước đầu ra |
| Ảnh | - | - | - | - | 640 × 640 |
| Conv0 | ReLU | 32 | 3 × 3/2 | 1 | 320 × 320 |
| Conv1 | ReLU | 64 | 3 × 3/2 | 1 | 160 × 160 |
| C2F2 | ReLU | 64 | 3 × 3/2 | 2 | 160 × 160 |
| Conv3 | ReLU | 128 | 3 × 3/2 | 1 | 80 × 80 |
| C2f4 | ReLU | 128 | 3 × 3/2 | 4 | 80 × 80 |
| Conv5 | ReLU | 256 | 3 × 3/2 | 1 | 40 × 40 |
| C2f6 | ReLU | 256 | 3 × 3/2 | 6 | 40 × 40 |
| Conv7 | ReLU | 512 | 3 × 3/2 | 1 | 20 × 20 |
| C2f8 | ReLU | 512 | 3 × 3/2 | 8 | 20 × 20 |
| SPPF | ReLU | 512 | - | 1 | 20 × 20 |
| Upsample10 | - | - | - | 1 | 40 × 40 |
| Mèo11 | ReLU | 512 | - | 1 | 40 × 40 |
| C2f12 | ReLU | 256 | 3 × 3 | 12 | 40 × 40 |
| Upsample13 | - | - | - | 1 | 80 × 80 |
| Concat14 | ReLU | 256 | - | 1 | 80 × 80 |
| C2F15 | ReLU | 128 | 3 × 3 | 15 | 80 × 80 |
| Conv16 | ReLU | 128 | 3 × 3 | 1 | 80 × 80 |
| Mèo17 | ReLU | 128 | - | 1 | 80 × 80 |
| C2F18 | ReLU | 64 | 3 × 3 | 18 | 80 × 80 |
| Conv19 | ReLU | 64 | 3 × 3 | 1 | 80 × 80 |
| Concat20 | ReLU | 64 | - | 1 | 80 × 80 |
| C2F21 | ReLU | 32 | 3 × 3 | 21 | 80 × 80 |
| Phát hiện22 | - | - | - | - | 80 × 80 |

YOLOv8 sử dụng các hàm sigma, đánh giá sự hiện diện của một đối tượng trong hộp giới hạn và các hàm softmax, phân loại đối tượng thành lớp tương ứng của nó. Không giống như các phiên bản YOLO trước đây, bằng cách kết hợp Complete Intersection over Union và chức năng tổn thất tiêu cự phân phối, các cải tiến mô hình nâng cao độ chính xác trong việc dự đoán hộp giới hạn và giải quyết việc phát hiện các mô hình nhỏ hơn. Entropy chéo nhị phân được sử dụng để tính toán tổn thất phân loại, do đó tăng cường phân loại nhiều lớp.

YOLOv8x cho thấy hiệu suất tốt hơn trên bộ dữ liệu, MS COCO test-dev 2007, đạt được độ chính xác trung bình 53,9% cho hình ảnh 640 pixel hoạt động ở tốc độ 280 khung hình / giây khi sử dụng TensorRT.

## Kiến trúc YOLOv10

YOLOv10 thiết lập một chuẩn mực mới bằng cách tăng tốc độ suy luận, loại bỏ triệt tiêu không tối đa và giới thiệu hệ thống gán nhãn kép [41]. Nó cân bằng giữa giám sát và suy luận nhanh chóng thông qua các nhánh một-nhiều và một-một, và nâng cao hiệu quả của các nhiệm vụ đào tạo và suy luận bằng cách áp dụng một kỹ thuật kết hợp tiêu chuẩn để duy trì tính đồng nhất trong 2 nhánh đó.

Mô hình YOLOv10 phân tích hình ảnh bằng cách sử dụng một tập hợp các lớp tích chập, Conv0, Conv1, C2f2 và C2f4, sử dụng kích hoạt ReLU, giảm kích thước đầu ra từ 320 × 320 xuống 80 × 80 pixel. Các lớp khác, cụ thể là SCDown5, C2f6 và C2f8, giảm kích thước pixel xuống 20 × 20. Các lớp Spatial Pyramid Pooling và Pyramid Squeeze Attention tinh chỉnh các tính năng. Sau đó, các lớp lấy mẫu tăng kích thước lên 80 × 80. Kiến trúc chi tiết của YOLOv10 với các lớp, bộ lọc và đầu ra được hiển thị trong Bảng 3.

Bảng 3: Kiến trúc YOLOv10.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Lớp | Kích hoạt | Bộ lọc | Kích thước | Lặp lại | Kích thước đầu ra |
| Ảnh | - | - | - | - | 640 × 640 |
| Conv0 | ReLU | 32 | 3 × 3/2 | 1 | 320 × 320 |
| Conv1 | ReLU | 64 | 3 × 3/2 | 1 | 160 × 160 |
| C2F2 | ReLU | 64 | 3 × 3/2 | 2 | 160 × 160 |
| Conv3 | ReLU | 128 | 3 × 3/2 | 1 | 80 × 80 |
| C2f4 | ReLU | 128 | 3 × 3/2 | 4 | 80 × 80 |
| SCDown5 | ReLU | 128 | 3 × 3/2 | 1 | 80 × 80 |
| C2f6 | ReLU | 256 | 3 × 3/2 | 6 | 40 × 40 |
| SCDown7 | ReLU | 256 | 3 × 3/2 | 1 | 40 × 40 |
| C2f8 | ReLU | 512 | 3 × 3/2 | 8 | 20 × 20 |
| SPPF9 | ReLU | 512 | - | 1 | 20 × 20 |
| PSA10 | ReLU | 512 | - | 1 | 20 × 20 |
| Upsample11 | - | - | - | 1 | 40 × 40 |
| Concat12 | ReLU | 512 | - | 1 | 40 × 40 |
| C2F13 | ReLU | 256 | 3 × 3 | 13 | 40 × 40 |
| Upsample14 | - | - | - | 1 | 80 × 80 |
| Concat15 | ReLU | 256 | - | 1 | 80 × 80 |
| C2F16 | ReLU | 128 | 3 × 3 | 16 | 80 × 80 |
| Conv17 | ReLU | 128 | 3 × 3 | 1 | 80 × 80 |
| Concat18 | ReLU | 128 | - | 1 | 80 × 80 |
| C2f19 | ReLU | 64 | 3 × 3 | 19 | 80 × 80 |
| SCDown20 | ReLU | 64 | 3 × 3 | 1 | 80 × 80 |
| Concat21 | ReLU | 64 | - | 1 | 80 × 80 |
| C2fCIB22 | ReLU | 32 | 3 × 3 | CIB | 80 × 80 |
| v10Phát hiện23 | - | - | - | - | 80 × 80 |

Với các lớp, nghĩa là tự chú ý một phần (PSA) và bị hạn chế về không gian (SC), kiến trúc này cải thiện độ chính xác và hiệu suất, đồng thời bao gồm độ sâu, giảm độ phân giải và theo chiều kim đồng hồ, tăng kích thước của kênh, kết cấu và gán nhãn kép để đào tạo và tối ưu hóa suy luận, do đó cho phép mô hình YOLOv10 xử lý việc phát hiện các vật thể bị tắc hoặc nhỏ với tốc độ và độ chính xác tốt hơn.

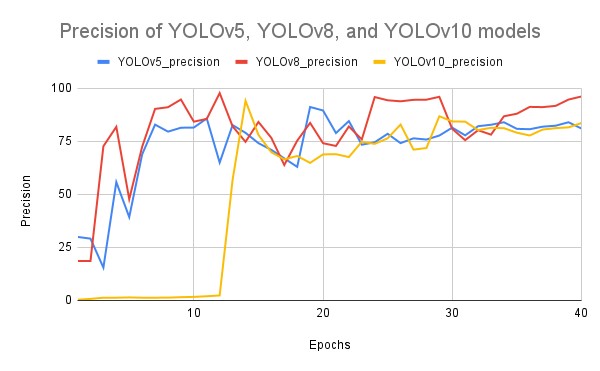
Mô hình YOLOv10 được cấu trúc theo cách nâng cao hiệu quả mà không ảnh hưởng đến hiệu suất bằng cách sử dụng đầu phân loại nhẹ để giải quyết vấn đề nút cổ chai liên quan đến các tác vụ hồi quy. Độ chính xác khoảng 70% đã đạt được

trong khi tích hợp tích chập theo chiều sâu và thêm 30% trong khi cùng được thêm vào một khối đảo ngược nhỏ gọn, kết hợp thêm với thiết kế khối hướng dẫn xếp hạng.

# Kết quả thí nghiệm

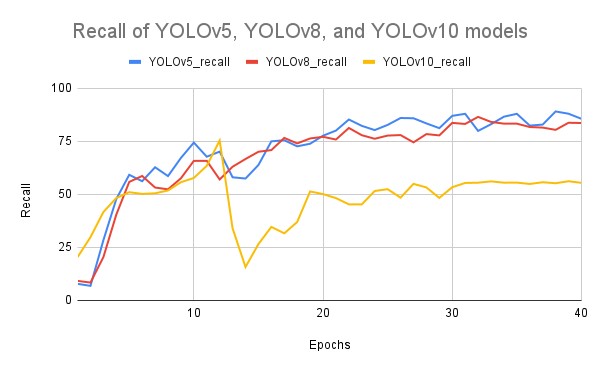
Các chỉ số hiệu suất của YOLOv5, YOLOv8 và YOLOv10 trong cả giai đoạn đào tạo và xác thực được thảo luận trong phần này. Việc đào tạo và xác nhận cả ba mô hình được thực hiện trong môi trường PyTorch với GPU NVIDIA, một hệ thống có hiệu suất cao. Để cải thiện độ chính xác và hiệu suất của các mô hình, các siêu tham số đã được điều chỉnh trong suốt 40 kỷ nguyên. Khóa đào tạo sử dụng trình tối ưu hóa AdamW với tốc độ học tập 0,001 và động lượng 0,9. Ngoài ra, các tham số nhóm được tinh chỉnh theo cách không nhìn thấy sự phân rã ở một số trọng lượng, trong khi các sai lệch nhận được sự chuẩn hóa L2 với tốc độ phân rã là 0,0005, ngăn ngừa sự phù hợp quá mức trong khi vẫn giữ được khả năng khái quát hóa.

Từ Hình 2, khi bộ đào tạo được đào tạo đến 40 kỷ nguyên, có thể lưu ý rằng các đường cong chính xác đào tạo cho cả 3 mô hình hội tụ sau 13 kỷ nguyên. YOLOv5 được cải thiện đáng kể, ổn định từ 60% đến 80%. YOLOv8 cho thấy độ chính xác cao nhất, tăng và trì trệ vào khoảng kỷ nguyên 7. Mặc dù YOLOv10 ban đầu hoạt động không tốt, nhưng sau 13 kỷ nguyên, nó đã tăng khoảng 0,7 và dao động khoảng 60% và 80%, cho thấy sự ổn định kém hơn. Cuối cùng, cần lưu ý rằng YOLOv8 có độ chính xác đào tạo cao nhất trong suốt 40 kỷ nguyên khi so sánh với các mô hình khác.



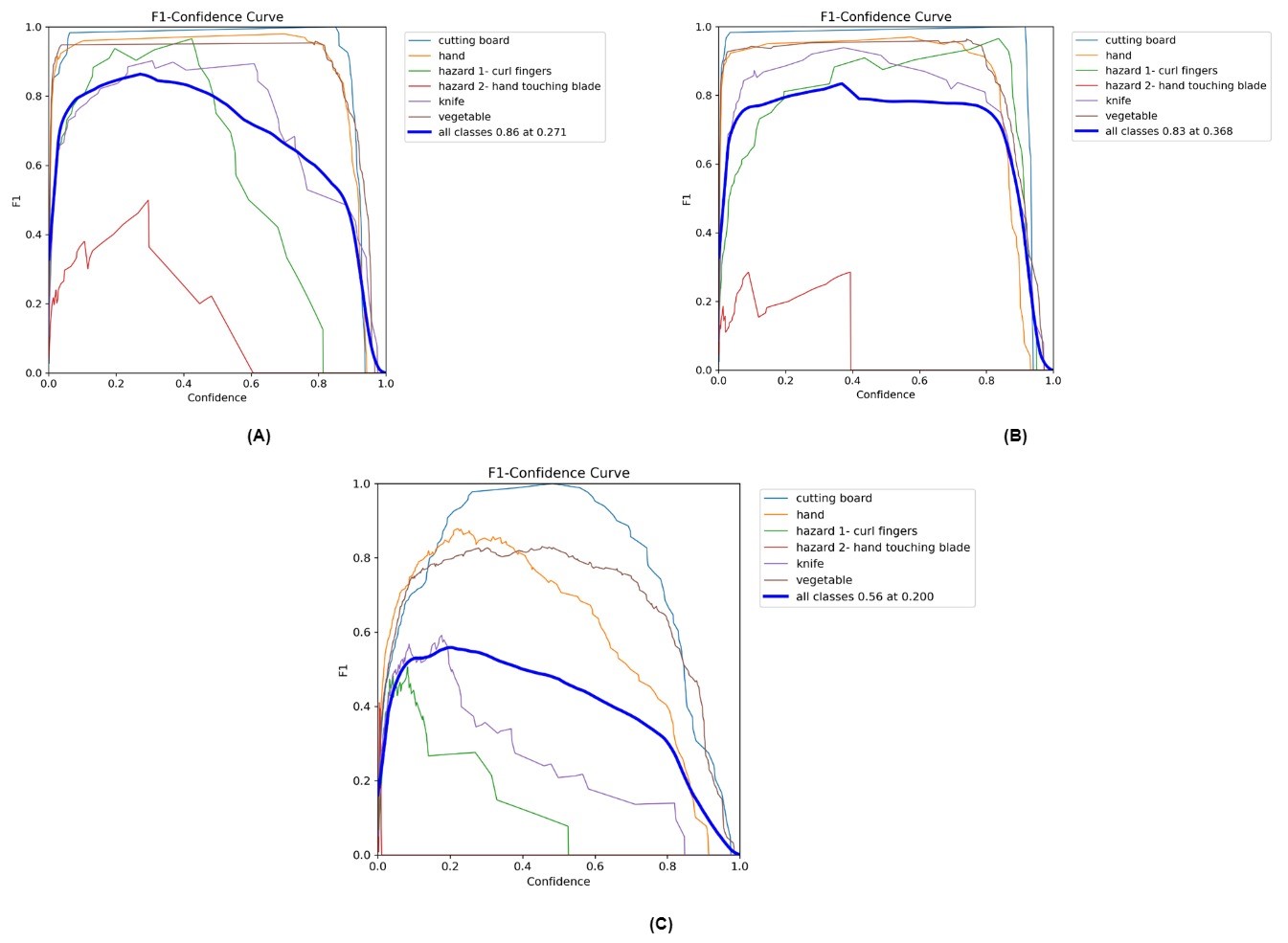
Hình 2: So sánh các giá trị chính xác trên ba mô hình.

Như thể hiện trong Hình 3, biểu đồ thu hồi so sánh YOLOv5, YOLOv8 và YOLOv10 qua 40 kỷ nguyên. Cũng giống như độ chính xác, YOLOv5 được cải thiện đều đặn và ổn định từ 60% đến 80% với một số biến động. YOLOv8 cho thấy hiệu suất tốt nhất, tăng khoảng 80%. YOLOv10 minh họa sự thay đổi, tăng ở mức 7, và sau đó giảm liên tục, cho thấy hiệu suất kém ổn định hơn của mô hình.



Hình 3: So sánh các giá trị thu hồi trên ba mô hình.

Các đường cong tin cậy F1 của mô hình thể hiện sự cân bằng giữa độ chính xác và khả năng thu hồi ở các mức độ tin cậy khác nhau, giải quyết việc xử lý dao an toàn với các lớp như thớt, tay, nguy hiểm 1: ngón tay cong, nguy hiểm 2: dao chạm tay, dao và rau. Các đường cong tin cậy F1 cho tất cả các mô hình với các lớp khác nhau được minh họa trong Hình 4A, 4B, 4C.



Hình 4: So sánh F1 so với sự tự tin trên ba mô hình.

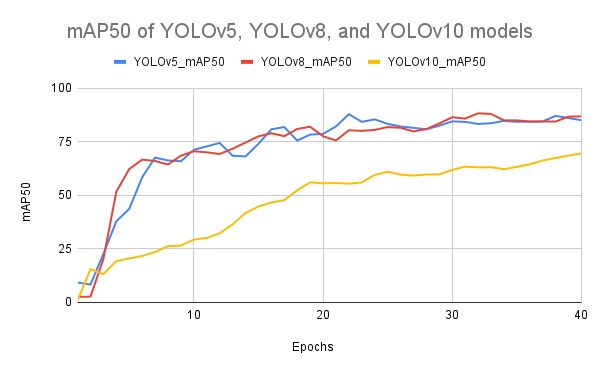
Đối với mô hình YOLOv5, các lớp, cụ thể là thớt, tay, dao và rau, có điểm F1 cao với các mức độ tin cậy khác nhau. Ngược lại, các hạng nguy hiểm (ngón tay cong và lưỡi chạm tay) có điểm F1 thấp hơn, đặc biệt là nguy hiểm 2 khi hiệu suất giảm với mức độ tự tin ngày càng tăng.

Điểm F1 kết hợp của tất cả các hạng đạt đỉnh 0,86 ở mức tin cậy 0,271 cho YOLOv5, do đó cung cấp sự cân bằng giữa độ chính xác và khả năng nhớ lại. Điều đó có nghĩa là tại thời điểm này, sẽ có tối thiểu dương tính giả hoặc âm tính với độ chính xác tốt hơn. Tuy nhiên, YOLOv8 có điểm F1 đỉnh thấp hơn một chút là 0,83 ở khoảng tin cậy là 0,368 và YOLOv10 có điểm F1 đỉnh thấp nhất là 0,56 ở mức 0,200. Những kết quả này chỉ ra rằng YOLOv5 vượt trội hơn hai mô hình còn lại.

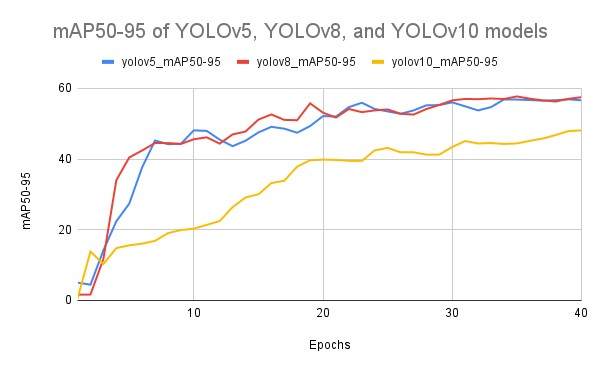
Trong YOLOv5, sau tỷ lệ 0,4 F1, sự tự tin của tất cả các lớp đã tăng lên 0,3 và sau đó giảm dần, trong khi sự tự tin của tất cả các lớp dao động ở mức 0,4 và sau đó bắt đầu giảm dần trong YOLOv8. Sự tự tin của tất cả các lớp trong YOLOv10 đã tăng từ 0,1 lên 0,2 và sau đó giảm sau đó.

Từ tất cả các biểu đồ, rõ ràng là tất cả các lớp học đều tăng lên khi mức độ tin cậy tăng lên đến một điểm nhất định và sau đó giảm sau một số mức độ tin cậy nhất định. Trong YOLOv5 và YOLOv8, lớp nguy hiểm 2 có tỷ lệ F1 thấp nhất với mức độ tin cậy thấp hơn, trong khi ở YOLOv10, lớp nguy hiểm 1 có tỷ lệ F1 và mức độ tin cậy thấp nhất. Trong cả ba mô hình, lớp thớt có tỷ lệ F1 cao nhất.

Ban đầu, YOLOv8 đạt 50% cho mAP50 (Hình 5) và 30% cho mAP50-95 (Hình 6) vào kỷ nguyên 5. YOLOv5 theo YOLOv8 cho đến kỷ nguyên 20, dao động trong khoảng 70% đến 80% đối với mAP50 và từ 50% đến 60% đối với mAP50-95. YOLOv10 được cải thiện với tốc độ ổn định; đến kỷ nguyên 40, nó đạt được 60% cho mAP50 và 45% cho mAP50-95. Nhìn chung, YOLOv8 là mô hình hiệu quả cho hiệu suất mAP tổng thể.



Hình 5: So sánh các giá trị mAP50 trên ba mô hình.



Hình 6: So sánh các giá trị mAP50-95 trên ba mô hình.

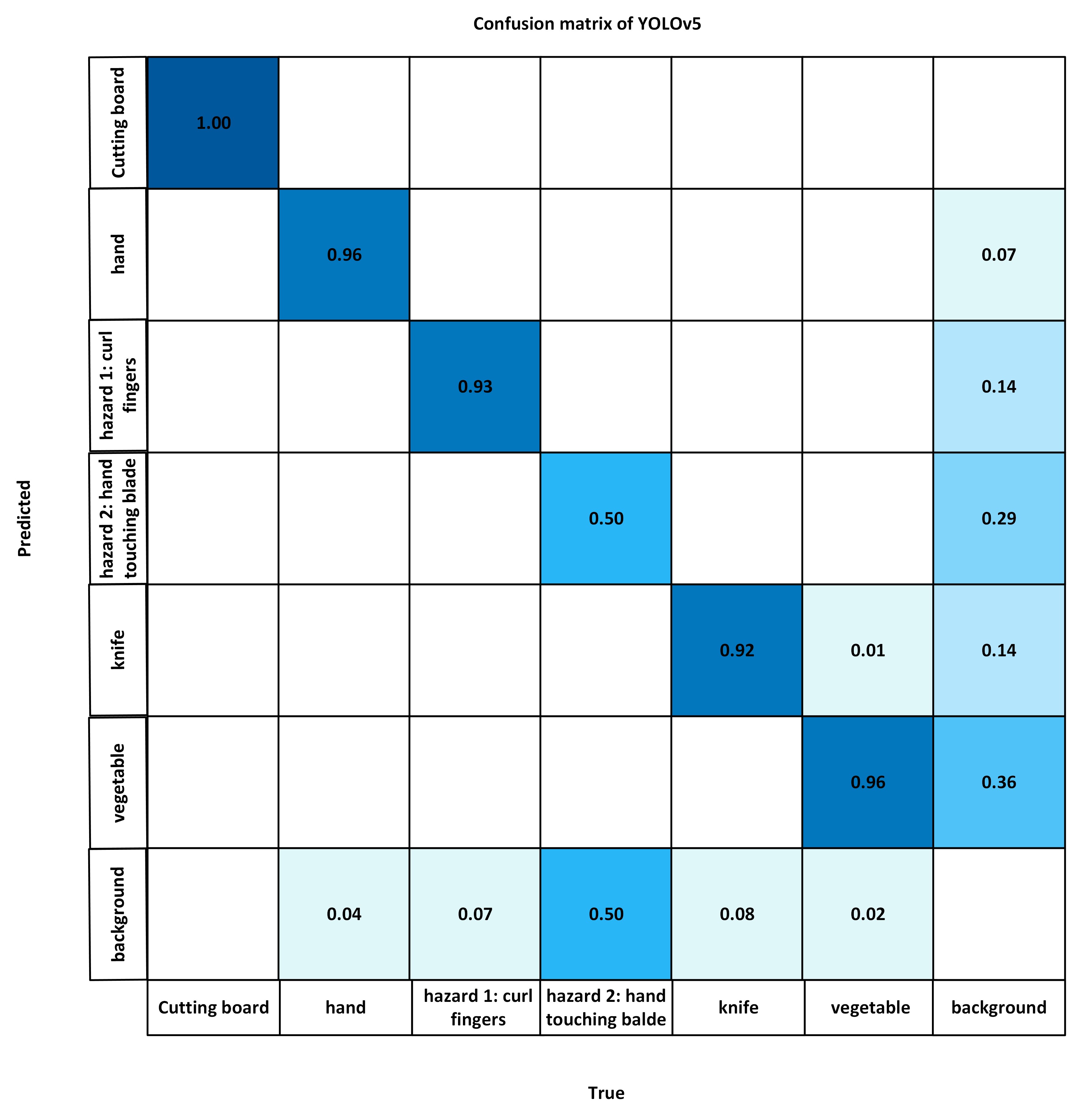
Độ chính xác phân loại cho tất cả các lớp trong ba mô hình, YOLOv5, YOLOv8 và YOLOv10, có thể được cung cấp bởi ma trận nhầm lẫn chuẩn hóa. Trong cả ba mô hình, lớp thớt được phân loại chính xác với độ chính xác cao nhất là 100%, cho thấy các mô hình có thể xác định hiệu quả danh mục mà không có lỗi.

YOLOv5 và YOLOv8 phân loại tay chính xác với độ chính xác cao là 0,96, trong khi điểm phân loại của mô hình YOLOv10 là 0,88, biểu thị một số thách thức trong việc xác định chính xác lớp này.

Lớp nguy hiểm 1 đã chứng minh sự khác biệt về hiệu suất của các mô hình. YOLOv8 hoạt động tốt với độ chính xác 1,00. Ngược lại, YOLOv5 và YOLOv10 cho thấy sự phân loại sai đáng kể với độ chính xác lần lượt là 0,50 và 0,16. Những điều này chỉ ra rằng mô hình YOLOv8 có khả năng nhận biết mối nguy hiểm 1 tốt hơn

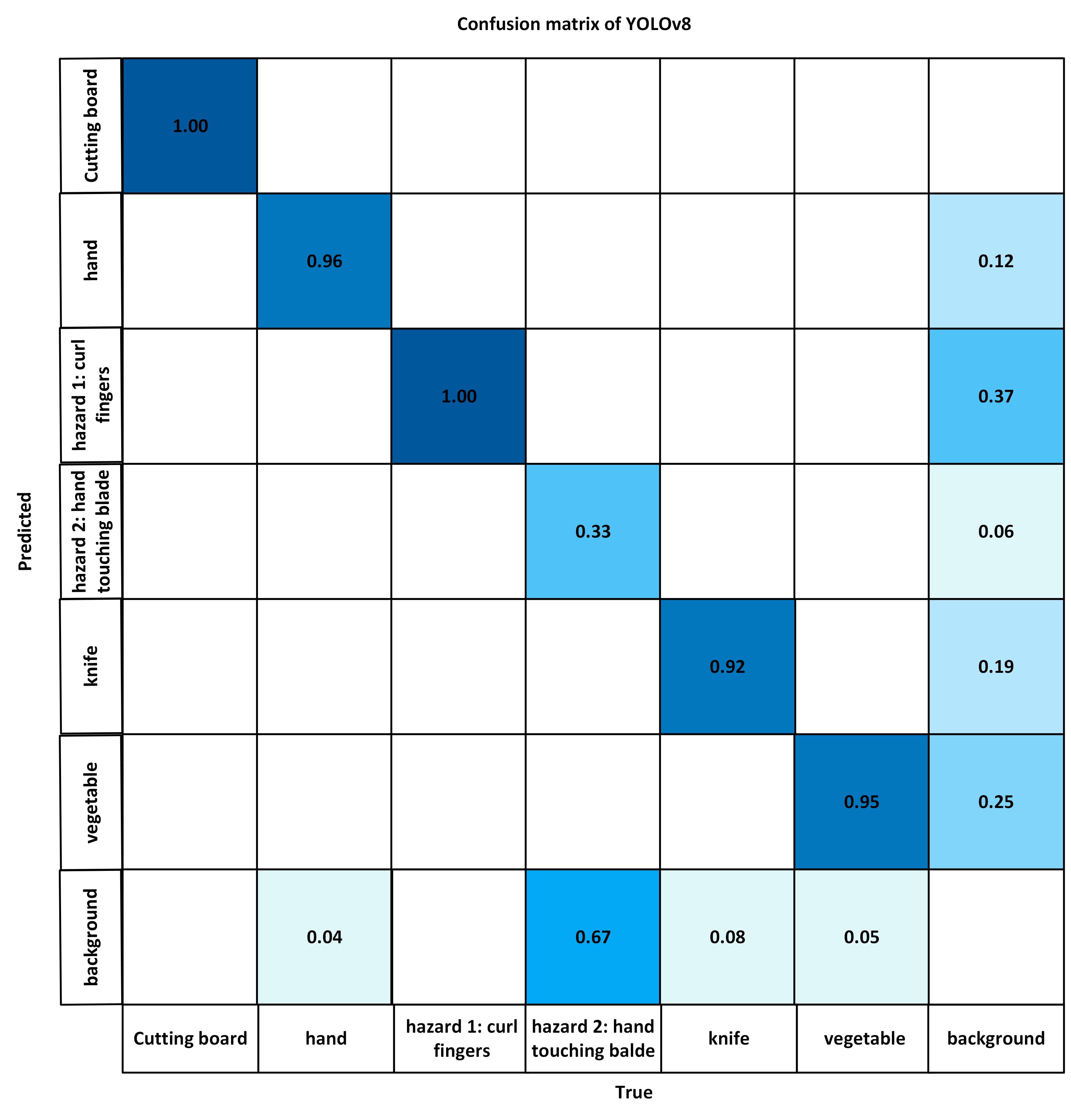
Không giống như các danh mục khác, tất cả các mô hình đều hoạt động kém cho danh mục nguy hiểm 2 với độ chính xác lần lượt là 0,50, 0,33 và 0,27 đối với YOLOv5, YOLOv8 và YOLOv10. Điều này có nghĩa là rất khó để tất cả các mô hình nhận ra mối nguy hiểm 2.

Trong danh mục dao, YOLOv5 (Hình 7) và YOLOv8 hoạt động tương tự, với độ chính xác 0,92; mặt khác, YOLOv10 thấp hơn một chút ở mức 0,89. Rau cho thấy độ chính xác cao trong YOLOv5 và YOLOv8, với 0,96 và 0,95, tương ứng, trong khi YOLOv10 giảm xuống 0,89.

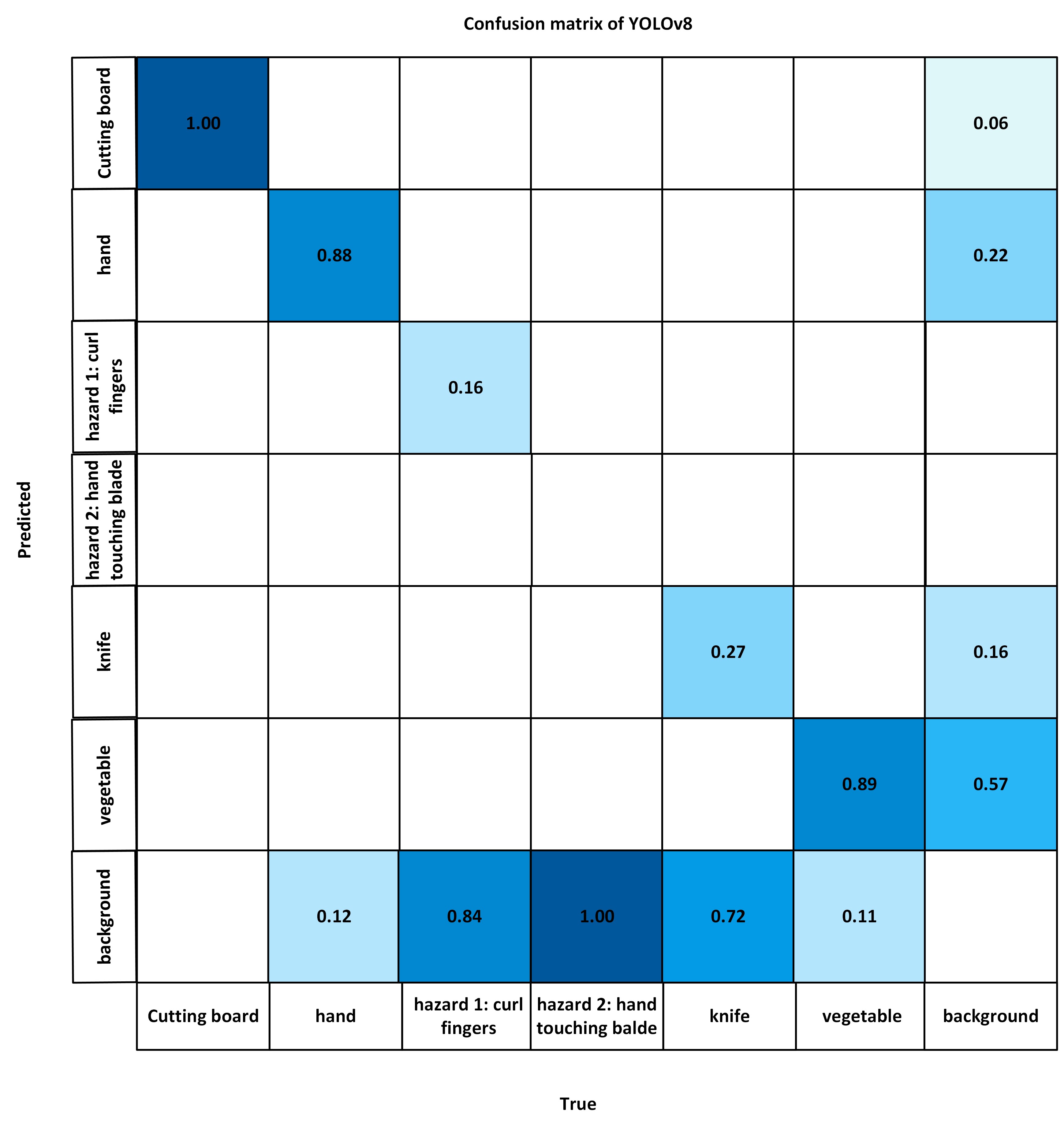


Hình 7: Ma trận nhầm lẫn chuẩn hóa YOLOv5.

Cuối cùng, nền tảng cho thấy các vấn đề phân loại sai trong tất cả các mô hình. YOLOv5 có độ chính xác 0,50, trong khi YOLOv8 (Hình 8 và YOLOv10 cho thấy tỷ lệ phân loại sai cao hơn với độ chính xác hơn 0,50 (lần lượt là 0,67 và 1,00). Điều này làm rõ rằng việc phân biệt các yếu tố nền vẫn là một thách thức, với YOLOv10 (Hình 9 là tồi tệ nhất.



Hình 8: Ma trận nhầm lẫn chuẩn hóa YOLOv8.



Hình 9: Ma trận nhầm lẫn chuẩn hóa của YOLOv10.

Tóm lại, mặc dù YOLOv8 có hiệu suất tốt nhất trong khi phân loại mối nguy hiểm 1, nhưng nó phải đối mặt với những thách thức trong khi nhận ra mối nguy hiểm 2 và nền. Ngoài ra, YOLOv10 có tỷ lệ phân loại sai cao nhất. Do đó, YOLOv5 dường như vượt trội hơn các mô hình khác.

# Sự thảo luận

Nghiên cứu này đề cập đến việc phân tích so sánh các mô hình YOLO, YOLOv5, YOLOv8 và YOLOv10, để xử lý dao an toàn trong nhà bếp. YOLOv8 đã chứng minh hiệu suất tốt hơn về độ chính xác tổng thể và khả năng thu hồi, đạt được độ chính xác cao một cách nhanh chóng trong quá trình đào tạo. Mô hình này cũng xuất sắc trong việc phân loại mối nguy hiểm 1 với độ chính xác 1,00 so với YOLOv5 chính xác là 0,50 và YOLOv10 là 0,16. Điều này là do cải tiến kiến trúc của mô hình YOLOv8, trích xuất tính năng tiên tiến và xử lý hiệu quả hơn các hình học đối tượng có khả năng xác định các đặc điểm của mối nguy hiểm 1. Ngoài ra, YOLOv8 đã thêm các lớp mới giúp cải thiện khả năng của mô hình để xác định từ dữ liệu đào tạo và áp dụng nó vào các tình huống trong thế giới thực.

Tuy nhiên, YOLOv5 cho thấy hiệu suất cân bằng giữa các hạng khác nhau, đạt được điểm số F1 kết hợp cao nhất. Mô hình này thực hiện tốt trong việc phân loại các lớp thớt, tay, dao và rau với điểm F1 cao ở các mức độ tin cậy khác nhau. Khả năng duy trì độ chính xác cân bằng và khả năng nhớ lại trên nhiều lớp của YOLOv5 đã làm nổi bật độ tin cậy của nó trong các tình huống khác nhau. Điều này có thể được liên kết với kiến trúc ít phức tạp hơn của nó, phát hiện các vật thể lớn hơn.

Trong khi phát hiện các lớp như thớt và tay, với độ chính xác lần lượt là 1,00 và 0,96, cả YOLOv5 và YOLOv8 đều hoạt động tốt, biểu thị rằng một trong hai mô hình có thể có hiệu quả trong việc xác định các lớp này. Những kết quả này cho thấy YOLOv8 linh hoạt hơn để phát hiện các đối tượng phức tạp, trong trường hợp này là mối nguy hiểm 1, trong khi YOLOv5 có thể được sử dụng trong các ứng dụng trong đó hiệu suất nhất quán trong các lớp khác nhau là điều cần thiết. YOLOv10 dường như không hoạt động

tốt.

Mặc dù nghiên cứu này cung cấp phân tích so sánh chi tiết về các mô hình YOLO (YOLOv5, YOLOv8 và YOLOv10) về an toàn dao trong môi trường nhà bếp, nhưng có một số lĩnh vực chưa được khám phá, chẳng hạn như môi trường nhà bếp khác nhau cùng với sự lộn xộn trên bàn hoặc các loại thực phẩm hoặc dao xử lý khác nhau. Trong tương lai, nghiên cứu này có thể được mở rộng hơn nữa để có các bộ dữ liệu lớn với các mối nguy hiểm trong khi xử lý dao.

Tóm lại, việc lựa chọn giữa YOLOv5, YOLOv8 và YOLOv10 liên quan đến việc đánh giá các đặc tính hiệu suất của chúng và kết hợp khả năng của mô hình với các yêu cầu của nhiệm vụ phát hiện.

# Kết luận

Hiệu suất của YOLOv5, YOLOv8 và YOLOv10 trong môi trường phát hiện an toàn dao đã được so sánh trong nghiên cứu này với ưu và nhược điểm của tất cả các mô hình cho năm lớp. Các phát hiện cho thấy YOLOv5 vượt trội hơn

YOLOv8 và YOLOv10 trong việc phát hiện xem bàn tay có tiếp xúc với lưỡi dao hay không, giải quyết mối nguy hiểm 1. Ngược lại, YOLOv8 vượt trội trong việc nhận biết các ngón tay có cuộn tròn trong tay cầm đồ hay không, phân tích mối nguy hiểm 2. YOLOv5, YOLOv8 và YOLOv10 chứng minh tỷ lệ chính xác gần như tương tự nhau trong việc phát hiện thớt. Từ nghiên cứu này, rõ ràng là các mô hình YOLOv5 và YOLOv8 có thể phát hiện mối nguy hiểm 2 và nguy cơ 1, tương ứng, mang lại ý nghĩa cho việc lựa chọn các mô hình phù hợp trên cơ sở các nhiệm vụ phát hiện.

Các công việc trong tương lai có thể liên quan đến việc tăng cường đánh giá để đánh giá hiệu suất của các mô hình trong việc sử dụng các loại thực phẩm và dao khác nhau và làm việc trong môi trường nơi các vật thể bị che khuất một phần hoặc toàn bộ bởi các dụng cụ và thiết bị. Ngoài ra, kích thước của tập dữ liệu có thể được tăng lên để bao gồm các mối nguy hiểm và các lớp khác. Nghiên cứu này cũng có thể được mở rộng sang các lĩnh vực khác như năng lượng tái tạo [42, 43] và chăm sóc sức khỏe [44].

# Tham khảo

1. Natthariya Laopracha và Khamron Sunat. Nghiên cứu so sánh về thời gian tính toán mà các tính năng dựa trên lợn được sử dụng để phát hiện xe. Trong *Những tiến bộ gần đây trong Công nghệ Thông tin và Truyền thông 2017: Kỷ yếu của Hội nghị Quốc tế lần thứ 13 về Máy tính và Công nghệ Thông tin (IC2IT),* trang 275–284. Mùa xuân, 2018.
2. J Kittler, M Hatef, Robert PW Duin và J Matas. Về việc kết hợp phân loại. Giao dịch IEEE về phân tích mẫu và trí tuệ máy móc, 1998.
3. H Wang, A Klaser, C Schmid và CL Liu. Nhận dạng hành động bằng quỹ đạo dày đặc. CVPR'11, Washington, DC, Hoa Kỳ. *IEEE Computer Society*, trang 3169–3176, 2011.
4. Marcus Rohrbach, Sikandar Amin, Mykhaylo Andriluka và Bernt Schiele. Một cơ sở dữ liệu để phát hiện hoạt động hạt mịn của các hoạt động nấu ăn. Trong *hội nghị IEEE năm 2012 về thị giác máy tính và nhận dạng mẫu*, trang 1194–1201. IEEE, 2012.
5. M Milagro Fernandez-Carrobles, Oscar Deniz và Fernando Maroto. Phát hiện súng và dao dựa trên r-cnn nhanh hơn để giám sát video. Trong *hội nghị Iberia về nhận dạng mẫu và phân tích hình ảnh*, trang 441–452. Mùa xuân, 2019.
6. Alex Krizhevsky, Ilya Sutskever và Geoffrey E Hinton. Phân loại imagenet với mạng nơ-ron tích chập sâu. *Truyền thông của ACM*, 60(6):84–90, 2017.
7. Huanhhuan Ran, Ship Wen, Kaibo Shi và Tingwen Huang. Thiết kế ổn định và nhỏ gọn của mạng nơ-ron googlenet memristive. *Máy tính thần kinh*, 441:52–63, năm 2021.
8. Kaiming He, Xiangyu Zhang, Shaoqing Ren và Jian Sun. Học tập dư sâu để nhận dạng hình ảnh. Trong *Kỷ yếu của hội nghị IEEE về thị giác máy tính và nhận dạng mẫu*, trang 770–778, 2016.
9. Tử Hàn Dương. Phân loại phong cách nghệ thuật hình ảnh dựa trên vggnet. Trong *Tạp chí Vật lý: Chuỗi hội nghị*, tập 1774, trang 012043. Nhà xuất bản IOP, 2021.
10. Ross Girshick, Jeff Donahue, Trevor Darrell và Jitendra Malik. Hệ thống phân cấp tính năng phong phú để phát hiện đối tượng chính xác và phân đoạn ngữ nghĩa. Trong *Kỷ yếu của hội nghị IEEE về thị giác máy tính và nhận dạng mẫu*, trang 580–587, 2014.
11. Mounika Gajja. Phát hiện khối u não bằng mặt nạ r-cnn. *J. Adv. Res. Dyn. Hệ thống điều khiển*, 12:101–108, 2020.
12. Shuang Liu, Xing Cui, Jiayi Li, Hui Yang và Niko Lukac. Phát hiện người đi bộ dựa trên r-cnn nhanh hơn.ˇ *Tạp chí quốc tế về kỹ thuật hiệu suất*, 15(7):1792, 2019.
13. Muhammad Hussain. Thuật toán lựa chọn biến thể Yolo-v5 cùng với sự gia tăng đại diện để mô hình hóa phương sai dựa trên sản xuất trong kiểm tra kệ pallet nhẹ tự động. *Dữ liệu lớn và điện toán nhận thức*, 7(2):120, 2023.
14. Joseph Redmon, Santosh Divvala, Ross Girshick và Ali Farhadi. Bạn chỉ nhìn một lần: Hợp nhất, phát hiện đối tượng theo thời gian thực. Trong *Kỷ yếu của hội nghị IEEE về thị giác máy tính và nhận dạng mẫu*, trang 779–788, 2016.
15. Joseph Redmon và Ali Farhadi. Yolo9000: tốt hơn, nhanh hơn, mạnh hơn. Trong *Kỷ yếu của hội nghị IEEE về thị giác máy tính và nhận dạng mẫu*, trang 7263–7271, 2017.
16. Joseph Redmon và Ali Farhadi. Yolov3: Một cải tiến gia tăng. *arXiv in trước arXiv:1804.02767*, 2018.
17. Alexey Bochkovskiy, Chien-Yao Wang và Hong-Yuan Mark Liao. Yolov4: Tốc độ tối ưu và độ chính xác của việc phát hiện đối tượng. *arXiv in trước arXiv:2004.10934*, 2020.
18. Lakshantha Dissanayake, Glenn Jocher, Q Burhan và Sergiu Waxmann. Hướng dẫn toàn diện về ultralytics yolov5, 2023.
19. Chuyi Li, Lulu Li, Hongliang Jiang, Kaiheng Weng, Yifei Geng, Liang Li, Zaidan Ke, Qingyuan Li, Meng Cheng, Weiqiang Nie, et al. Yolov6: Một khung phát hiện đối tượng một giai đoạn cho các ứng dụng công nghiệp. *arXiv in trước arXiv:2209.02976*, 2022.
20. Chien-Yao Wang, Alexey Bochkovskiy và Hong-Yuan Mark Liao. Yolov7: Túi miễn phí có thể đào tạo thiết lập nhà nước-of-the-nghệ thuật mới cho các máy dò đối tượng thời gian thực. Trong *Kỷ yếu của hội nghị IEEE / CVF về thị giác máy tính và nhận dạng mẫu*, trang 7464–7475, 2023.
21. Glenn Jocher, Muhammad Rizwan Munawar và Ayush Chaurasia. Yolo: Tóm tắt lịch sử, năm 2023.
22. Mujadd Al Rabbani Alif, Sabbir Ahmed và Muhammad Abul Hasan. Nhận dạng ký tự viết tay bangla bị cô lập với mạng nơ-ron tích chập. Năm *2017, Hội nghị quốc tế lần thứ 20 về máy tính và công nghệ thông tin (ICCIT),* trang 1–6. IEEE, 2017.
23. Cường Phạm và Patrick Olivier. Slice &dice: Nhận biết các hoạt động chuẩn bị thức ăn bằng gia tốc kế nhúng. Trong *Hội nghị châu Âu về tình báo xung quanh*, trang 34–43. Mùa xuân, 2009.
24. Sebastian Stein và Stephen J McKenna. Kết hợp gia tốc kế nhúng với thị giác máy tính để nhận biết các hoạt động chuẩn bị thực phẩm. Trong *Kỷ yếu của hội nghị chung quốc tế ACM 2013 về điện toán phổ biến và phổ biến*, các trang 729–738, 2013.
25. Marek Zywicki, Andrzej Matiola ̇ nski, Tomasz M Orzechowski và Andrzej Dziech. Phát hiện dao như một tập hợp con của phương pháp phát hiện đối tượng dựa trên thác haar. Trong *Kỷ yếu Hội nghị quốc tế lần thứ 11 "Nhận dạng mẫu và xử lý thông tin*, các trang 139–142, 2011.
26. Alex Krizhevsky, Ilya Sutskever và Geoffrey E Hinton. Phân loại imagenet với mạng nơ-ron tích chập sâu. *Truyền thông của ACM*, 60(6):84–90, 2017.
27. Songbo Chen, Wenhu Tang, Tianyao Ji, Huiling Zhu, Ye Ouyang và Wenbo Wang. Phát hiện đội mũ bảo hiểm an toàn dựa trên r-cnn được cải thiện nhanh hơn. Năm *2020 Hội nghị chung quốc tế về mạng nơ-ron (IJCNN)*, trang 1–7. IEEE, 2020.
28. David A Noever và Sam E Miller Noever. Máy dò dao và mối đe dọa. *arXiv in trước arXiv:2004.03366*, 2020.
29. Samet Akcay, Mikolaj E Kundegorski, Chris G Willcocks và Toby P Breckon. Sử dụng kiến trúc mạng nơ-ron tích chập sâu để phân loại và phát hiện đối tượng trong hình ảnh an ninh hành lý bằng tia X. *Giao dịch IEEE về pháp y và bảo mật thông tin*, 13(9):2203–2215, 2018.
30. Lê Văn Hùng, Hải Vũ và Nguyễn Thị Thúy. Một công việc khung hỗ trợ người khiếm thị: phát hiện đối tượng phổ biến và ước tính tư thế trong môi trường xung quanh. Năm *2018, hội nghị NAFOSTED lần thứ 5 về thông tin và khoa học máy tính (NICS),* trang 216–221. IEEE, 2018.
31. Joshua van Staden và Dane Brown. Đánh giá các thuật toán dựa trên yolo để phát hiện tay trong nhà bếp. Trong

*Hội nghị quốc tế năm 2021 về Trí tuệ nhân tạo, Dữ liệu lớn, Hệ thống điện toán và truyền thông dữ liệu (icABCD),* trang 1–7. IEEE, năm 2021.

1. Qiang Li, Feng Zhao, Zhongping Xu, Kexin Li, Jing Wang, Haofeng Liu, Liang Qin và Kaipei Liu. Cải tiến thuật toán yolov4 để quản lý an toàn công việc hệ thống điện tại chỗ. *Báo cáo năng lượng*, 8: 739–746, 2022.
2. Hubert Ngankam, Philippe Dion, Hélène Pigot và Sylvain Giroux. Theo dõi nhiều đối tượng theo thời gian thực cho các hoạt động nấu ăn an toàn. Trong *Hội nghị Quốc tế về Nhà thông minh và Viễn thông Y tế*, trang 192–204. Mùa xuân, năm 2023.
3. Yunfan Shi, Zheng Yang, Yifei Bi, Jingcheng Li, Xiaohui Zhu và Yong Yue. Phát hiện mặt nạ thời gian thực của nhân viên nhà bếp bằng yolov5 và điện toán biên. Năm *2023, Hội nghị quốc tế lần thứ 3 về Máy tính, Điều khiển và Robot (ICCCR),* trang 33–40. IEEE, năm 2023.
4. Isaias Majil, Mau-Tsuen Yang và Sophia Yang. Hướng dẫn nấu ăn tương tác dựa trên thực tế tăng cường. *Cảm biến*, 22(21):8290, 2022.
5. Iker Azurmendi, Ekaitz Zulueta, Jose Manuel Lopez-Guede, Jon Azkarate và Manuel González. Cảm biến mặt bếp dựa trên thuật toán phát hiện đối tượng yolo. *Cảm biến*, 23(5):2780, 2023.
6. Saydirasulov Norkobil Saydirasulovich, Akmalbek Abdusalomov, Muhammad Kafeel Jamil, Rashid Nasimov, Dinara Kozhamzharova, và Young-im Cho. Phương pháp phát hiện cháy cải tiến dựa trên yolov6 cho môi trường thành phố thông minh. *Cảm biến*, 23(6):3161, 2023.
7. Yi Yi Aung và Kyi Zar Oo. Phát hiện hình ảnh súng và dao dựa trên yolo v7. Năm *2024, Hội nghị quốc tế lần thứ 3 về trí tuệ nhân tạo cho Internet vạn vật (AIIoT),* trang 1–6. IEEE, năm 2024.
8. Maha Mokrani và Zied Hajaiej. Phát hiện đối tượng thời gian thực với sự thay đổi dữ liệu. *Przeglad Elektrotechniczny*, 2024(5), 2024.
9. Anuja Radhakrishnan, Sumisha Samuel, Sachin Shaju John, Riya Ann Reji, Stephin John, Liya Elizabeth Jacob và Devi Vinod. Một hệ thống cảnh báo tự động để giám sát vệ sinh trong nhà hàng bằng cách sử dụng thị giác máy. Năm *2024, Hội nghị quốc tế lần thứ 1 về xu hướng trong các hệ thống và công nghệ kỹ thuật (ICTEST),* trang 1–6. IEEE, năm 2024.
10. Mujadd Al Rabbani Alif và Muhammad Hussain. Yolov1 đến yolov10: Đánh giá toàn diện về các biến thể yolo và ứng dụng của chúng trong lĩnh vực nông nghiệp. *arXiv in trước arXiv:2406.10139*, 2024.
11. Arsalan Zahid, Muhammad Hussain, Richard Hill và Hussain Al-Aqrabi. Mạng tích chập nhẹ để phát hiện khuyết tật quang điện tự động. Năm *2023, Hội nghị quốc tế lần thứ 9 về xu hướng công nghệ thông tin (ITT),* trang 133–138. IEEE, năm 2023.
12. Muhammad Hussain, Mahmoud Dhimish, Violeta Holmes và Peter Mather. Triển khai mạng rbf dựa trên AI cho quy trình phát hiện lỗi quang điện. *AIMS Điện tử và Kỹ thuật điện*, 4 (1): 1–18, 2019.
13. Muhammad Hussain. Khi nào, ở đâu và cái nào?: Điều hướng giao điểm của thị giác máy tính và AI tạo ra để tích hợp kinh doanh chiến lược. *Truy cập IEEE*, 11:127202–127215, 2023.