

# Nhận diện thiết bị bảo hộ cá nhân (Personal Protective Equipment – PPE) trong thi công xây dựng

Phan Văn Đạt  
phandat12082002@gmail.com

Ninh Thị Hương  
(VTX)

## 1. Tóm tắt

Nhận diện thiết bị bảo hộ cá nhân là một việc rất cần thiết trong các hoạt động công nghiệp như xây dựng, khai thác khoáng sản, chế tạo,... để đảm bảo an toàn lao động cho công nhân tránh khỏi các sự cố đáng tiếc. Vì vậy, trong nghiên cứu này, chúng tôi đề xuất ứng dụng mô hình YOLOv7 để phát hiện việc thiếu các thiết bị bảo hộ của công nhân trước cũng như trong quá trình thực hiện công việc thi công xây dựng. Mô hình đề xuất đã đạt kết quả 89.9% với độ đo mAP@0.5 tính trung bình trên các lớp với thời gian suy diễn trung bình là 20ms/frame trên GPU T4 của Google Colab.

## 2. Giới thiệu

Thiết bị bảo hộ lao động là những dụng cụ, phương tiện cần thiết mà người lao động phải được trang bị để sử dụng trong khi làm việc hoặc thực hiện nhiệm vụ để bảo vệ cơ thể khỏi tác động của các yếu tố nguy hiểm, độc hại phát sinh trong quá trình lao động, khi các giải pháp công nghệ, thiết bị, kỹ thuật an toàn, vệ sinh lao động tại nơi làm việc chưa thể loại trừ hết. Tuy nhiên, thực trạng hiện nay, công nhân không ý thức được nguy hiểm khi làm việc mà bỏ qua việc trang bị đầy đủ thiết bị bảo hộ, dẫn đến các sự cố đáng tiếc xảy ra. Nguyên nhân còn có thể bắt nguồn từ những người sử dụng lao động khi không có hoặc ít phân công người hướng dẫn hoặc giám sát lao động hoặc thiếu quy định khi người lao động bắt đầu làm việc. Để giảm thiểu tình trạng này, ứng dụng các mô hình học sâu để tự động hóa việc giám sát trang bị bảo hộ của công nhân là một việc cần thiết và hiệu quả để vừa đảm bảo an toàn cho công nhân, đảm bảo đúng quy định khi sử dụng lao động, vừa tiết kiệm chi phí cho người sử dụng lao động.

Trong nghiên cứu này, chúng tôi đề xuất sử dụng mô hình YOLOv7 huấn luyện trên tập dữ liệu cá nhân để phát hiện chính xác việc trang bị đầy đủ thiết bị bảo hộ lao động của công nhân trong quá trình thi công công trình xây dựng bao gồm: mũ bảo hiểm, áo phản quang và găng tay.



Hình 1. Mô tả bài toán

Sau đây, chúng tôi xin trình bày một số hoạt động nghiên cứu đi trước liên quan đến đề tài, mô tả về bộ dữ liệu, phương pháp và kết quả thực hiện.

## 3. Những hoạt động liên quan

Nghiên cứu về phát hiện thiết bị bảo hộ cá nhân trước đây như “Nipun et al.[1] Deep learning for site safety: Real-time detection of personal protective equipment”, nghiên cứu sử dụng mô hình YOLOv3 để phát hiện áo phản quang và mũ bảo hiểm của công nhân và đã đạt được kết quả khá tốt. Một số nghiên cứu tương tự như “Venkata et al.[2] Detection of Personal Protective Equipment (PPE) Compliance on Construction Site Using Computer Vision Based Deep Learning Techniques” cũng đạt được kết quả khả quan.

## 4. Dữ liệu

Bộ dữ liệu sử dụng để huấn luyện mô hình được tập hợp từ các bộ dữ liệu về PPE trên trang web Roboflow. Bộ dữ liệu bao gồm 670 ảnh được chia thành các tập huấn luyện, đánh giá và kiểm tra theo tỉ lệ 8:1:1. Các ảnh được gán bounding box theo yêu cầu bài toán với 7 lớp: *Person*, *Safety Vest*, *Helmet*, *Glove*, *NO-Safety Vest*, *NO-Helmet*, *NO-Glove*.

#### 4.1. Tùy chỉnh bounding box

Do các ảnh được tập hợp từ nhiều nguồn khác nhau, các nhãn có sẵn trong ảnh chưa phù hợp với yêu cầu bài toán, chúng tôi sử dụng công cụ roboflow annotation để vẽ lại bounding box và gán nhãn cho bounding box theo 7 lớp như trên.

Lớp	Số lượng
Person	1102
Glove	775
Helmet	661
Safety Vest	603
NO-Glove	531
NO-Safety Vest	455
NO-Helmet	453

Bảng 1. Số lượng bounding box của từng lớp trên toàn bộ dữ liệu sau khi chỉnh sửa

#### 4.2. Chuẩn hóa dữ liệu

Để phù hợp cho việc huấn luyện mô hình YOLOv7, chúng tôi tiến hành định lại kích thước tất cả các ảnh về kích thước 640x640. Dữ liệu được chia thành các thư mục theo định dạng của YOLOv7.

```
dataset/
├── train/
│   ├── images/
│   │   ├── img1.jpg
│   │   └── ...
│   └── labels/
│       ├── img1.txt
│       └── ...
├── valid/
│   ├── images/
│   │   ├── img2.jpg
│   │   └── ...
│   └── labels/
│       ├── img2.txt
│       └── ...
├── test/
│   ├── images/
│   │   ├── img3.jpg
│   │   └── ...
│   └── labels/
│       ├── img3.txt
│       └── ...
└── data.yaml
```

Hình 2. Định dạng thư mục dữ liệu YOLOv7

#### 4.3. Data augmentation

Tập các ảnh huấn luyện được làm giàu bằng các phương pháp:

- Xoay ngẫu nhiên ảnh sang trái hoặc phải  $90^0$
- Xoay ngẫu nhiên ảnh một góc từ  $-15^0$  đến  $15^0$

#### 5. Phương pháp

##### 5.1. Mô hình YOLOv7 [3]

Mô hình YOLOv7 được đánh giá là có tốc độ vượt trội và độ chính xác vẫn ổn định so với các mô hình nhận diện vật thể hiện tại nhờ kiến trúc được thêm vào một số sửa đổi bao gồm:

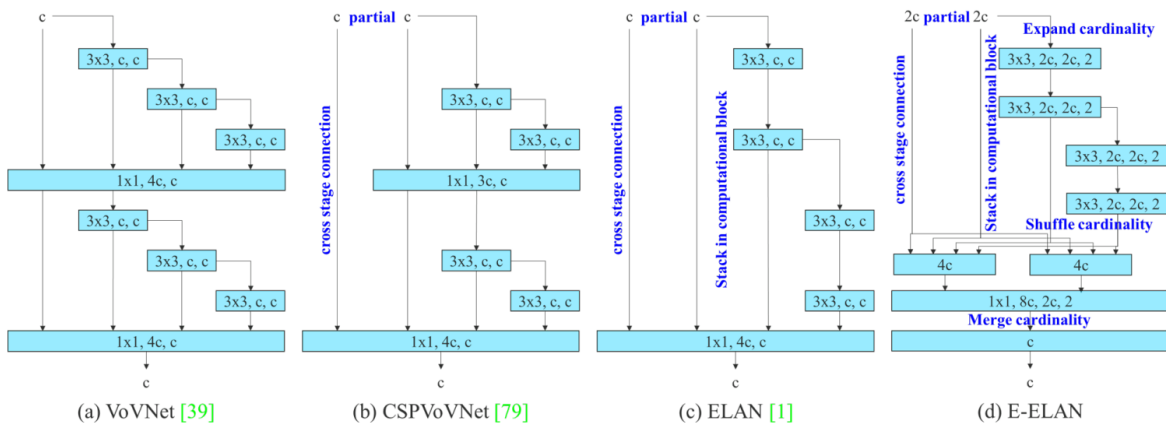
- E-ELAN (Extended Efficient Layer Aggregation Network)
- Model Scaling for Concatenation-based Models
- Trainable Bag-of-freebies (BoF)
  - Planned re-parameterized convolution
  - Coarse for auxiliary and Fine for lead loss

##### 5.1.1. Mạng E-ELAN

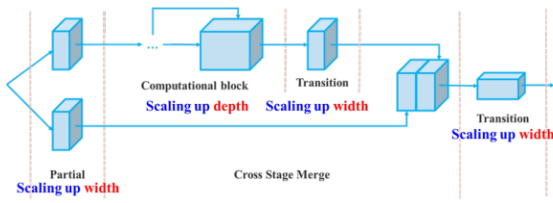
E-ELAN là một khối tính toán của YOLOv7 backbone, sử dụng tích chập nhóm để tăng số lượng features và tổng hợp lại bằng cách xáo trộn sau đó gộp lại mà không thay đổi đường truyền gradient ban đầu. Cách này giúp tăng khả năng học của mô hình và tối ưu hơn việc tính toán và sử dụng tham số.

##### 5.1.2. Model Scaling for Concatenation-based Models

Mục đích của model scaling là để điều chỉnh một số thuộc tính của mô hình, đưa vào nhiều thang đo khác nhau nhằm đáp ứng những yêu cầu tốc độ suy diễn khác nhau.



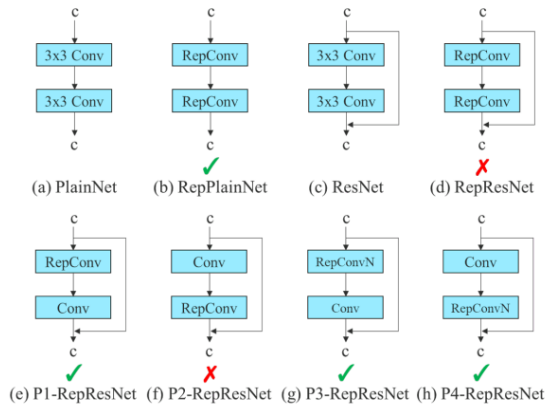
Hình 3. E-ELAN và một số mô hình trước về cải thiện hiệu quả của tầng



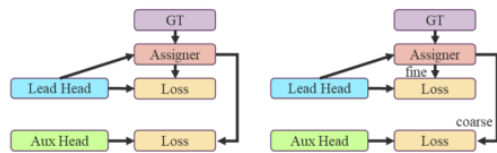
Hình 4. Compound scaling up depth and width for concatenation-based model

### 5.1.3. Trainable Bag-of-freebies (BoF)

BoF là một phương pháp tăng hiệu năng của mô hình mà không làm tăng chi phí huấn luyện. Trong đó, kỹ thuật “planned re-parameterized convolution” giúp tăng tốc độ suy diễn của mô hình và “coarse for auxiliary and Fine for lead loss” giúp tăng độ chính xác của mô hình.



Hình 5. planned re-parameterized model



(d) Lead guided assigner (e) Coarse-to-fine lead guided assigner

Hình 6. lead guided assigner

## 5.2. Hàm loss

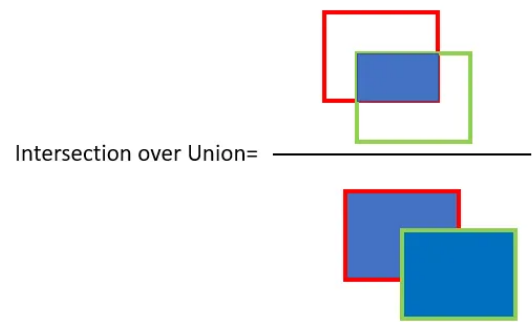
Hàm loss của YOLOv7 được tính từ tổng 3 hàm loss với trọng số riêng của mỗi hàm bao gồm:

- Objectness loss: là giá trị cross-entropy nhị phân giữa xác suất tồn tại vật thể trong bounding box với giá trị CIoU giữa bounding box đang xét và mục tiêu.

- Box loss: được tính bằng trung bình giá trị  $(1 - \text{CIoU})$  của toàn bộ anchor box với mục tiêu tương ứng
- Classification loss: là giá trị cross-entropy nhị phân giữa xác suất của các lớp với one-hot vector của lớp nhãn đúng

### 5.3. Các chỉ số đánh giá

IoU (Intersection over Union) là chỉ số thường dùng để đánh giá mô hình nhận diện vật thể, được tính là tỉ lệ giữa phần giao và phần hợp của 2 bounding box: mục tiêu và dự đoán.



Hình 7. Intersection over Union

Đặt ngưỡng cho IoU là  $\alpha$ , khi đó:

- Nếu  $\text{IoU} \geq \alpha$  thì kết quả được tính là True Positive (TP)
- Nếu  $\text{IoU} < \alpha$  thì kết quả được tính là False Positive (FP)
- Khi mô hình không nhận diện được vật thể, kết quả tính là False Negative (FN)

#### 5.3.1. Precision và Recall

Precision và Recall được tính theo TP, FP và FN theo công thức:

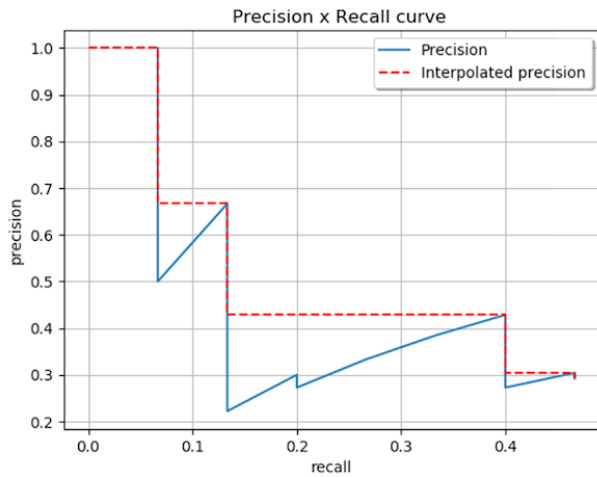
$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

Precision thể hiện độ chính xác của mô hình khi dự đoán là Positive. Recall thể hiện khả năng bao quát các mẫu Positive.

#### 5.3.2. Mean Average Precision (mAP)

Precision và recall của từng lớp sau khi được tính với các giá trị confidence khác nhau, được biểu diễn trên đường cong PR, AP của lớp được tính theo phương pháp “interpolation all point”. Giá trị mAP được tính là trung bình giá trị AP của tất cả các lớp.



Hình 8. Tính AP theo interpolation all point

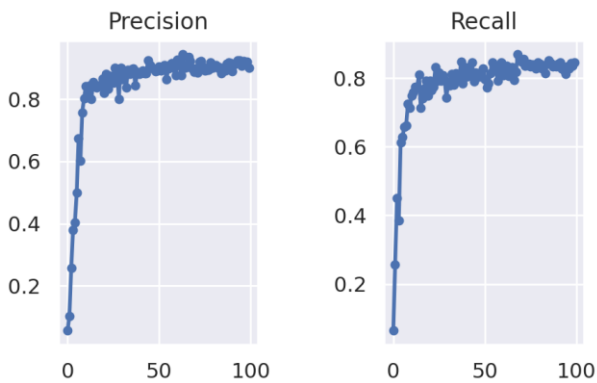
## 6. Thực nghiệm và kết quả

### 6.1. Tham số huấn luyện

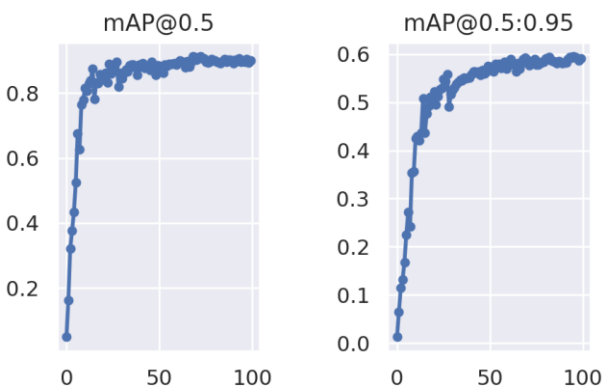
Mô hình được cài đặt và huấn luyện trên Google Colaboratory với GPU T4 với 100 epochs và cấu hình mặc định của YOLOv7 trong 2 giờ.

### 6.2. Kết quả

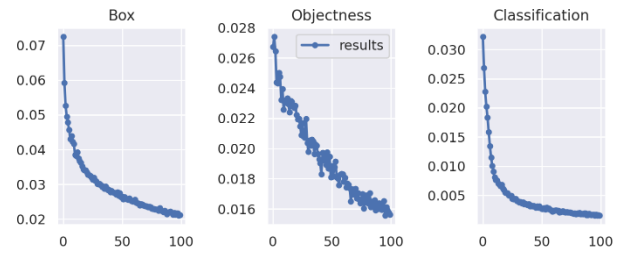
Mô hình được đánh giá dựa trên các độ đo precision, recall và mAP. Dưới đây là mô tả quá trình học của mô hình.



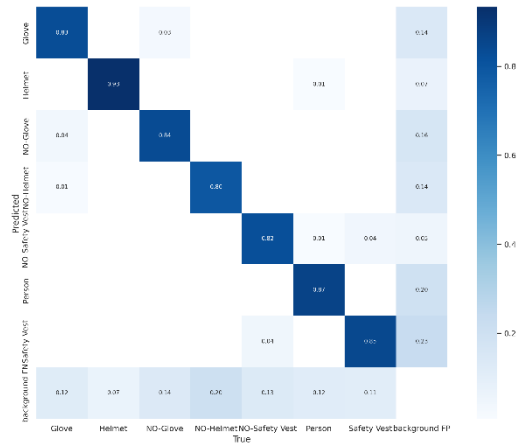
Hình 9. Precision và recall



Hình 10. mAP@0.5 và mAP@[0.5:0.95:0.05]

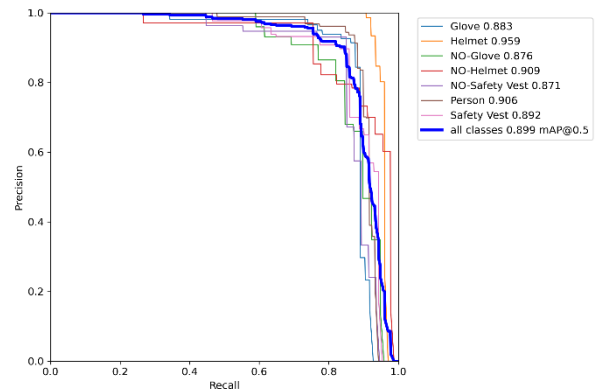


Hình 11. Các giá trị loss

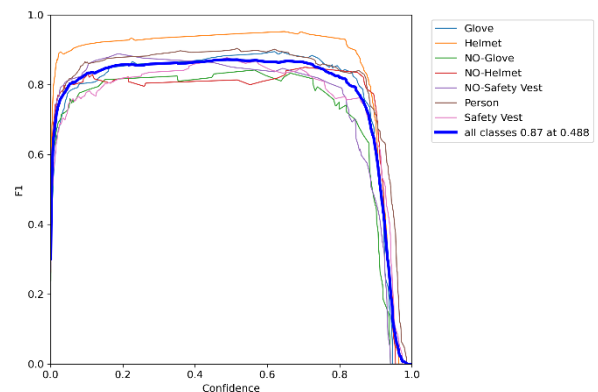


Hình 12. Confusion matrix

Sau khi huấn luyện, chúng tôi thu được đường cong precision-recall và giá trị mAP@0.5 của mô hình cho kết quả tốt nhất.



Hình 13. Đường cong PR



Hình 14. Chỉ số F1



Nhìn vào đồ thị kết quả huấn luyện có thể thấy, mô hình sau huấn luyện có kết quả nhận diện các lớp theo yêu cầu bài toán khá tốt. Dưới đây là một số hình ảnh kết quả của mô hình chạy trên tập kiểm tra.



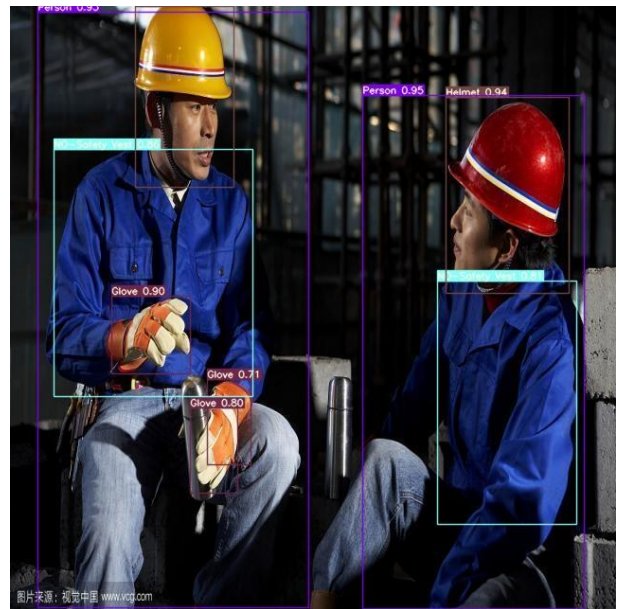
Hình 15. Một số hình ảnh kết quả (1)



Hình 16. Một số hình ảnh kết quả (2)

### 6.3. Phân tích lỗi

Sau khi chạy mô hình sau huấn luyện trên tập kiểm tra, chúng tôi nhận thấy mô hình còn một số hạn chế như sau: Vẫn còn trường hợp nhận diện nhầm hoặc thiếu vật thể; một số trường hợp một vật thể có 2 bounding box đè lên nhau. Dưới đây là minh họa kết quả lỗi.



Hình 17. 1 găng tay có 2 bounding box của lớp Glove



Hình 18. Nhận diện nhầm lớp Person

## 7. Kết luận

Trong nghiên cứu này, chúng tôi đã sử dụng ứng dụng của mô hình nhận diện vật thể YOLOv7 để phát hiện thiết bị bảo hộ cá nhân cho công nhân thi công xây dựng. Mô hình đã đạt được kết quả khá tốt tuy nhiên vẫn còn một số hạn chế nhất định. Do đó, sau nghiên cứu này, chúng tôi sẽ tiến hành thu thập thêm dữ liệu và có thể điều chỉnh lại một số tham số mô hình để cải thiện độ chính xác của mô hình.

## Tài liệu tham khảo

- [1] Nipun D. Nath, Amir H. Behzadan, Stephanie G. Paal (2020). Deep learning for site safety: Real-time detection of personal protective equipment.  
<https://www.sciencedirect.com/science/article/pii/S0926580519308325>
- [2] Venkata Santosh Kumar Delhi\*, R. Sankarlal and Albert Thomas (2020). Detection of Personal Protective Equipment (PPE) Compliance on Construction Site Using Computer Vision Based Deep Learning Techniques.  
<https://www.frontiersin.org/articles/10.3389/fbuil.2020.00136/full>
- [3] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors
- [4] [Online]  
<https://towardsdatascience.com/yolov7-a-deep-dive-into-the-current-state-of-the-art-for-object-detection-ce3ffedeeae3>