

Trash Detection

Samsung Innovation Campus

Nguyen Vi Chi Cuong Do Tien Dung Nguyen Tien Dung Vu Duc Duy

VNU Hanoi University of Science

Ngày 27 tháng 6 năm 2025

- ▶ Giới thiệu và đặt vấn đề
- ▶ Các phương pháp sử dụng
- ▶ Kết luận



- **Tình trạng môi trường hiện nay:**

- Sự phát triển khoa học kỹ thuật và nhu cầu sống tăng cao.
- Lượng rác thải gia tăng, gây ô nhiễm đất, nước, không khí.
- Ảnh hưởng tiêu cực đến sức khỏe cộng đồng và hệ sinh thái.

- **Giải pháp công nghệ:**

- Ứng dụng trí tuệ nhân tạo (AI) trong phát hiện và phân loại rác thải.
- Hệ thống thông minh giúp xử lý nhanh, chính xác.
- Nâng cao ý thức bảo vệ môi trường.

- **Phân loại rác thải:**

- Nhựa, kim loại, thủy tinh, chất hữu cơ, ...
- Nguồn gốc: sinh hoạt, công nghiệp, nông nghiệp.

- **Hậu quả nếu không xử lý đúng cách:**

- Ô nhiễm môi trường nghiêm trọng.
- Ảnh hưởng đến hệ sinh thái và sức khỏe con người.

- **Tầm quan trọng của phân loại tại nguồn:**

- Cơ sở cho quản lý rác thải hiện đại.
- Hỗ trợ tái chế và giảm thiểu chất thải.

- Xây dựng hệ thống nhận diện và phân loại rác thải tự động từ hình ảnh.
- Chức năng chính:
 - Người dùng tải ảnh lên hệ thống.
 - Sử dụng mô hình học sâu để xác định loại rác (nhựa, kim loại, giấy, ...).
 - Hiển thị kết quả trực quan: nhãn, độ tin cậy, vùng phát hiện.
 - Lưu trữ kết quả để phân tích và cải tiến.



- **Lợi ích thực tiễn:**

- Hỗ trợ phân loại rác tại nguồn.
- Nâng cao hiệu quả tái chế.
- Giảm ô nhiễm môi trường.

- **Đóng góp khoa học:**

- Thể hiện vai trò của AI trong giải quyết vấn đề môi trường.
- Góp phần xây dựng các giải pháp công nghệ bền vững.
- Bảo vệ hành tinh cho thế hệ tương lai.

- **Thực trạng:**

- Lượng rác thải ngày càng gia tăng.
- Công tác thu gom và phân loại chủ yếu thủ công, chưa tối ưu.
- Rác thải không được phân loại tại nguồn, gây khó khăn cho tái chế.

- **Hệ quả:**

- Giảm hiệu quả xử lý rác.
- Tồn thời gian và nguồn lực.
- Nguy cơ ô nhiễm và ảnh hưởng sức khỏe người lao động.

- **Nhu cầu:**

- Hệ thống phân loại rác thải tự động.
- Đảm bảo vệ sinh, nâng cao hiệu suất, bảo vệ môi trường.

- **Phương pháp:**

- Sử dụng thị giác máy tính và học sâu (deep learning).
- Áp dụng các mô hình: YOLOv11, Faster R-CNN, Rt-DETR.
- Nhận diện và phân loại rác thải qua hình ảnh từ camera hoặc thiết bị đầu vào.

1. Thu thập và gán nhãn dữ liệu ảnh:

- Thu thập hình ảnh rác thải.
- Gán nhãn (nhựa, kim loại, giấy, thực phẩm, ...).

2. Huấn luyện mô hình phân loại:

- Huấn luyện mô hình học sâu dựa trên đặc trưng hình ảnh.

3. Nhận diện và phân loại:

- Hệ thống tự động nhận diện loại rác thải từ ảnh mới.
- Đảm bảo phân loại nhanh chóng, chính xác.

4. Đưa ra quyết định xử lý:

- Đưa ra hướng xử lý: tái chế, tiêu hủy, thu gom riêng.



- **Hiệu quả:**

- Tăng hiệu suất phân loại rác thải.
- Giảm thiểu thời gian và nguồn lực.

- **Ứng dụng thực tiễn:**

- Triển khai tại đô thị, khu công nghiệp, hộ gia đình.
- Hỗ trợ quản lý rác thải thông minh.

- **Bảo vệ môi trường:**

- Góp phần xây dựng các giải pháp công nghệ bền vững.

Mục lục



- ▶ Giới thiệu và đặt vấn đề
- ▶ Các phương pháp sử dụng
- ▶ Kết luận





- **YOLO (You Only Look Once):**
 - Mô hình học sâu chuyên cho phát hiện đối tượng (object detection).
 - Cơ chế một bước (one-stage detection).
- **YOLOv11:**
 - Cải tiến độ chính xác, giữ tốc độ xử lý nhanh.
 - Phù hợp cho hệ thống thời gian thực như phân loại rác.
 - Tư tưởng: “Nhìn một lần – nhận diện tất cả” (one-shot prediction).

Các cải tiến của YOLOv11



- **Backbone hiện đại:**

- Kết hợp ConvNet và attention module (Transformer, EfficientNet).
- Tăng khả năng trích xuất đặc trưng hình ảnh từ nhiều cấp độ.

- **Neck module hiệu quả:**

- BiFPN (Bidirectional Feature Pyramid Network).
- Tổng hợp thông tin từ các tầng, cải thiện phát hiện vật thể nhỏ.

Các cải tiến của YOLOv11 (tt)



- **Detection head anchor-free:**

- Thay anchor boxes truyền thống bằng cơ chế anchor-free.
- Giảm độ phức tạp, tăng khả năng phát hiện vật thể không ổn định (túi nilon nhăn, giấy vo tròn, rác hữu cơ).

- **Tích hợp attention:**

- Sử dụng CBAM hoặc Self-Attention.
- Cải thiện nhận diện trong môi trường nhiễu nền, ánh sáng kém.

Phù hợp với phân loại rác thải



- **Lý do YOLOv11 phù hợp:**

- Xử lý hiệu quả các đối tượng nhỏ, hình dạng bất quy tắc.
- Đáp ứng tốt với góc chụp không đồng nhất.
- Tốc độ nhanh, phù hợp cho ứng dụng thời gian thực.

- **Ứng dụng thực tiễn:**

- Phân loại rác thải tự động từ hình ảnh.
- Hỗ trợ quản lý rác thải thông minh, bền vững.

Tổng quan về thuật toán Faster R-CNN



- **Faster R-CNN:**

- Mô hình phát hiện vật thể (object detection) hai giai đoạn.
- Kết hợp đề xuất vùng (region proposal) tự động và phân loại vật thể chính xác trong cùng một mạng CNN.

- **Thành phần chính:**

- Backbone Network (ResNet-50 + FPN)
- Region Proposal Network (RPN)
- ROI Pooling/ROI Align
- Detection Head (phân loại + tính chỉnh bounding box)

- **Ứng dụng:** Phát hiện vật thể nhỏ, chồng lấn, phù hợp bài toán rác thải.

Cấu hình huấn luyện Faster R-CNN



- **Bộ dữ liệu:**

- Nguồn: Roboflow, hơn 22,000 ảnh các loại rác thải.
- Tập huấn luyện: 19,824 ảnh (87.5%)
- Tập kiểm định: 1,935 ảnh (8.5%)
- Tập kiểm tra: 909 ảnh (4%)

- **Tham số huấn luyện:**

- Optimizer: AdamW
- Batch size: 16
- Learning rate: 0.000500979
- Input image size: 640×640 pixels

- **Backbone:** ResNet-50 + FPN

- **RPN:** 9 anchor boxes mỗi vị trí (3 tỷ lệ, 3 kích thước)

Quy trình huấn luyện và tối ưu tham số



- **Chuẩn bị dữ liệu:**
 - Resize và chuẩn hóa ảnh, chuyển đổi nhãn bounding box.
- **Khởi tạo mô hình:**
 - Sử dụng pre-trained ResNet-50-FPN, fine-tune trên tập rác thải.
- **Huấn luyện:**
 - Chia batch (16 ảnh), cập nhật trọng số bằng AdamW.
- **Đánh giá:**
 - Kiểm tra hiệu suất trên tập validation sau mỗi epoch.
- **Tối ưu tham số:**
 - Sử dụng Optuna để tối ưu learning rate, batch size.
 - Learning rate là tham số quan trọng nhất.



- **Chỉ số đánh giá:**

- mAP@0.5: 0.44164 (trial tối ưu thứ 3)
- Precision: 0.62
- Recall: 0.58
- Training loss: giảm từ 0.8 xuống 0.2 sau 20 epochs
- Validation loss: hội tụ ở mức 0.3, không overfitting rõ rệt

- **Chất lượng phát hiện:**

- Tốt với vật thể chồng lấn, góc chụp khác nhau.
- Khó khăn với vật thể quá nhỏ hoặc bị che khuất.

- **Tốc độ suy luận:** 5 FPS (chậm hơn YOLO)

- **So sánh YOLO11:** YOLO11 nhanh hơn (38.2 FPS), nhưng Faster R-CNN chính xác hơn với vật thể nhỏ/chồng lấn.



- **Ưu điểm:**

- Độ chính xác cao, đặc biệt với vật thể nhỏ và chồng lấn.
- Tùy chỉnh linh hoạt: thay đổi backbone, số anchor boxes.
- Ổn định trong huấn luyện, ít overfitting.

- **Hạn chế:**

- Tốc độ suy luận chậm (5 FPS).
- Yêu cầu phần cứng mạnh.
- Phụ thuộc nhiều vào tối ưu tham số, đặc biệt là learning rate.

- Faster R-CNN là mô hình mạnh cho phát hiện rác thải, đặc biệt khi cần độ chính xác cao.
- Tối ưu tham số và lựa chọn backbone phù hợp giúp cải thiện hiệu suất.
- Cần cân nhắc giữa độ chính xác và tốc độ khi triển khai thực tế.



- **RT-DETR (Real-Time Detection Transformer):**

- Mô hình phát hiện một bước (one-stage), kết hợp tốc độ và độ chính xác cao.
- Dựa trên kiến trúc DETR nhưng được cải tiến cho phù hợp thời gian thực.
- Học trực tiếp mối quan hệ không gian giữa các đối tượng bằng Transformer.

- **Đặc điểm nổi bật:**

- Thiết kế đơn giản, end-to-end, không cần anchor boxes hay NMS.
- Phù hợp với các ứng dụng phát hiện rác thải từ hình ảnh.

Các cải tiến trong RT-DETR



- **Decoder hiệu quả:**

- Sử dụng các kỹ thuật như Deformable Attention hoặc khởi tạo truy vấn hai giai đoạn.
- Tăng tốc độ hội tụ, nâng cao hiệu suất mô hình.

- **Backbone tối ưu:**

- Hỗ trợ các kiến trúc mạnh như ResNet, ConvNeXt, Swin Transformer.
- Cải thiện khả năng phát hiện đối tượng nhỏ, che khuất, ở xa.

Quy trình triển khai RT-DETR



- **Tiền xử lý dữ liệu:**

- Dữ liệu định dạng COCO với nhãn bounding box.
- Ảnh được resize, chuẩn hóa và chia thành các tập train/val/test.

- **Huấn luyện mô hình:**

- Mô hình được tinh chỉnh từ trọng số pretrained.
- Sử dụng tối ưu hóa hiện đại để huấn luyện toàn bộ pipeline end-to-end.

Tối ưu tham số RT-DETR



- **Chiến lược tối ưu:**

- Tối ưu siêu tham số như tốc độ học, trọng số phạt bằng các phương pháp hiện đại.
- Áp dụng kỹ thuật điều chỉnh phù hợp với dữ liệu thực tế.

- **Kết quả đạt được:**

- Mô hình hội tụ nhanh, ổn định.
- Cải thiện hiệu quả phát hiện đối tượng trong bài toán phân loại rác.



- **Ưu điểm:**

- Kết hợp tốc độ và độ chính xác trong một mô hình phát hiện hiện đại.
- Thiết kế đơn giản, không cần các bước xử lý trung gian phức tạp.

- **Ứng dụng thực tiễn:**

- Rất phù hợp cho hệ thống phân loại rác tự động có yêu cầu cao về độ chính xác.
- Có khả năng mở rộng và điều chỉnh linh hoạt theo yêu cầu sử dụng.

Bảng: So sánh mAP50-95 (%) trên các tập dữ liệu

Mô hình	Train	Validation	Test
Faster R-CNN	49.096	39.034	41.467
YOLOv11	70.868	54.640	58.047
RT-DETR	64.166	51.213	53.576

- YOLOv11 dẫn đầu cả 3 tập dữ liệu.
- Chênh lệch train \rightarrow test:
 - Faster R-CNN: -7.63%
 - RT-DETR: -10.59%
 - YOLOv11: -12.82%
- Hiệu suất tổng thể: YOLOv11 > RT-DETR > Faster R-CNN

Phân tích hiệu suất theo lớp (Tập Test)



Bảng: AP@[0.50:0.95] theo từng lớp vật thể (%)

Lớp	Faster R-CNN	YOLOv11	RT-DETR
Battery	54.546	79.435	72.401
Glass	39.070	57.529	54.384
Medical	35.918	52.909	48.794
Metal	31.124	45.496	45.567
Organic	24.058	40.515	32.665
Paper	33.100	51.776	47.013
Plastic	29.872	46.261	40.495
SmartPhone	84.048	90.457	90.291

- YOLOv11 dẫn đầu 6/8 lớp vật thể.
- Lớp khó nhất: Organic ($\leq 40.52\%$)
- Lớp tốt nhất: SmartPhone ($\geq 84.05\%$)

Đánh giá tốc độ và ổn định



Tốc độ suy luận

- YOLOv11: **31.5 ms/ảnh** (nhANH NHẤT)
- RT-DETR: 35.0 ms/ảnh
- Faster R-CNN: ~50 ms/ảnh

ĐỘ ỔN ĐỊNH

- **Biến động train \rightarrow validation:**
 - RT-DETR: -12.95% (ỔN ĐỊNH NHẤT)
 - YOLOv11: -16.23%
 - Faster R-CNN: -20.50%
- **Nguyên nhân:**
 - Faster R-CNN: bị overfitting
 - YOLOv11: cần nhiều dữ liệu hơn

Ưu điểm nổi bật

- **YOLOv11 vượt trội:**
 - mAP50-95 cao nhất (58.05%)
 - Tốc độ xử lý nhanh nhất (31.5 ms/ảnh)
 - Hiệu quả với vật thể nhỏ
- **RT-DETR:** ổn định nhất qua các tập dữ liệu

Hạn chế và khuyến nghị

- **Hạn chế:**
 - Hiệu suất thấp với lớp Organic và vật thể trung bình ($\leq 40.52\%$)
 - Biến động hiệu suất từ train sang test
- **Khuyến nghị:**
 - Sử dụng YOLOv11 cho ứng dụng real-time
 - Kết hợp FPN với RT-DETR để cải thiện vật thể nhỏ

Mục lục



- ▶ Giới thiệu và đặt vấn đề
- ▶ Các phương pháp sử dụng
- ▶ **Kết luận**





- Đã đánh giá ba mô hình phát hiện đối tượng nổi bật:
 - Faster R-CNN
 - YOLOv11
 - RT-DETR
- Mục tiêu: lựa chọn mô hình tối ưu cho bài toán phân loại rác thải.
- Đánh giá dựa trên: độ chính xác (mAP), tốc độ suy luận, khả năng tổng quát và hiệu suất theo từng lớp.



- **YOLOv11:**

- Độ chính xác và tốc độ vượt trội.
- Phù hợp nhất cho ứng dụng thời gian thực.

- **RT-DETR:**

- Hiệu suất ổn định, khả năng tổng quát hóa tốt.
- Có tiềm năng phát triển trong các hệ thống cần độ chính xác cao.

- **Faster R-CNN:**

- Cơ bản, nhưng kém hiệu quả hơn trong bài toán này.



- **Lớp Organic:**

- Mọi mô hình đều gặp khó khăn trong việc nhận diện.
- Nguyên nhân: tính đa dạng hình dạng, thiếu dữ liệu đại diện.

- **Vật thể kích thước trung bình:**

- Hiệu suất phát hiện chưa cao.
- Cần cải thiện khả năng biểu diễn đặc trưng đa tỷ lệ.



- **Tăng cường dữ liệu:**
 - Bổ sung dữ liệu lớp Organic.
 - Đa dạng hóa kích thước và điều kiện ảnh.
- **Tối ưu kiến trúc:**
 - Kết hợp các kỹ thuật FPN, PAN vào RT-DETR.
- **Kết hợp mô hình (Ensemble):**
 - Kết hợp YOLOv11 và RT-DETR để tận dụng ưu điểm của cả hai.



- YOLOv11 là mô hình phù hợp nhất cho hệ thống phân loại rác thông minh.
- RT-DETR là lựa chọn tiềm năng nếu cần độ chính xác cao và mở rộng.
- Nghiên cứu cung cấp nền tảng vững chắc cho các giải pháp AI ứng dụng vào quản lý môi trường và phân loại rác.