BÁO CÁO TECHNICAL TEST

Applicable solutions for
few-shot object detections for aerial images

CT UAV

Nguyễn Đức Quốc Bình

AI Techical Test

Department: Applied AI & Machine Learning

Question 1

What are the applicable solutions for few-shot object detection for aerial images? Please provide your solutions (with all scientific references), explaining how it works and why it is feasible.

Question 2

Implement a small demo for your solution in Question 1 (preferably in Pytorch). Please submit your Github link along with a technical report regarding your solution performance and the visualization of the result.

1. Mục tiêu

Xây dựng một giải pháp có thể phát hiện vật thể trong trường hợp chỉ có 1 – 5 ảnh mẫu (few-shot) cho các đối tượng cụ thể ứng dụng cho những bức ảnh trên không.

Giải pháp kết hợp YOLO – World, một model sử dụng open-vocabulary object detection, cho phép mô hình nhận dạnh vật thể thông qua các text prompt mà không cần huấn luyện lại.

2. Giải pháp đề xuất

2.1. YOLO-World

YOLO-World (2024) là mô hình open-vocabulary object detection.

- Được phát triển bởi Tencent AI và cộng động Ultralytics
- Dựa trên YOLOv8 được tích hợp thêm vision-language để kết nối hình ảnh với văn bản. Nhờ đó cho phép phát hiện đối tượng trong ảnh theo mô tả bằng từ khoá.

Điểm mạnh:

- Không cần huấn luyện lại
- Tốc độ nhanh
- Hỗ trợ cho ít mẫu dữ liệu

Hạn chế:

- Với những hình ảnh aerial nhiều vật thể, zero-accuracy có thể chưa cao, vẫn còn phát hiện thiếu một số đối tượng trong trường hợp nhiều vật thể.

2.2. Few-shot Feature Prototyping

Ý tưởng: Với mỗi lớp sẽ có một vài hình ảnh từ 1 – 5 ảnh để hỗ trợ phát hiện

Lấy mỗi lớp sẽ có 1 – 5 ảnh hỗ trợ. Sau đó sẽ lấy vector đặc trưng của mỗi ảnh và tính ra vector trung bình. Khi YOLO – World phát hiện bouding box, sẽ:

- Cắt vùng đó
- Rút những đặc trưng CLIP
- So sánh với vector trung bình đã có
- Nếu cosine similarity (thước đo giống nhau giữa 2 vector) cao -> giữ lại, ngược lại thấp thì sẽ loại ra

Việc kết hợp này sẽ tăng đáng kể độ chính xác

3. Quy trình hoạt động

3.1. Zero shot bằng YOLO-World

Khi này YOLO-World sẽ dự đoán các bounding box từ các text prompt đã được nhập ví dụ như: airplane, ship,…

3.2. Tạo các feature Prototype

Với mỗi lớp được nhập sẽ có các ảnh hỗ trợ để mô hình có thể nhận biết chính xác hơn. Sau đó sẽ tạo ra các vector đặc trưng để hỗ trợ việc tính toán.

Sau đó chạy để tạo vector prototype

3.3. Few-shot cosine similarity so sánh sự giống nhau 2 vector

Crop vùng phát hiện vật thể

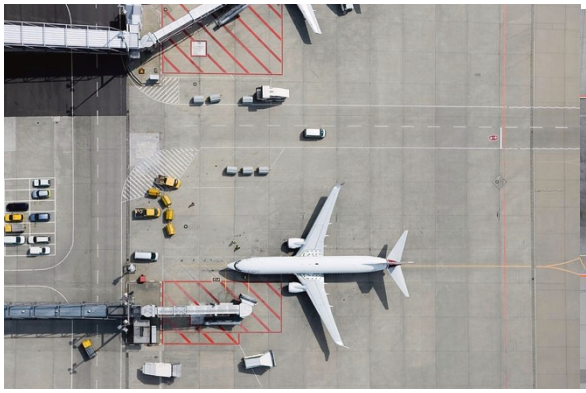
Rút các vector đặc trưng

Tính cosine similarity để so sánh xem 2 vector giữa đối tượng mẫu và đối tượng mới.

Kết quả cao thì sẽ giữ lại – ngược lại thấp thì sẽ loại bỏ

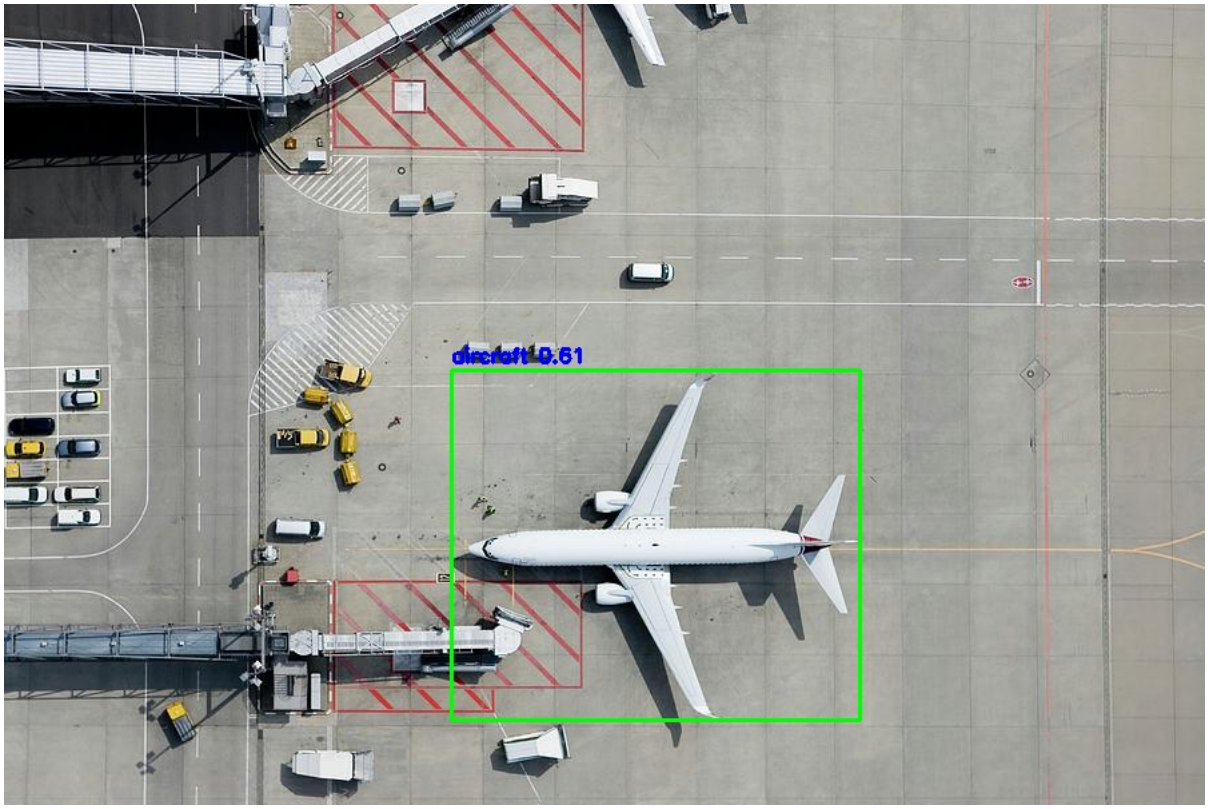
4. Kết quả Demo

4.1. Ảnh trước khi demo



dreamstime

4.2. Sau khi demo



5. Ưu & Nhược điểm

5.1. Ưu điểm

Không cần phải huấn luyện lại mô hình

Hoạt động tốt với việc ít dữ liệu

Tốc độ xử lý khá nhanh

Có thể mở rộng thêm các lớp mới chỉ với ảnh và từ khoá

5.2. Nhược điểm

Phụ thuộc độ phân giải của ảnh

Các lớp giống nhau có thể gây nhiễu

Ảnh với nhiều vật thể khi detect có thể bỏ qua một số vật thể

6. Source code

https://github.com/quocbinhn/CT-UAV-R-D---Technical-Test

7. Tài liệu tham khảo

https://docs.ultralytics.com

https://blog.roboflow.com/what-is-yolo-world

https://www.youtube.com/watch?v=7EmfDY2Ooqo

https://arxiv.org/abs/2103.00020

https://arxiv.org/abs/1503.03832

https://arxiv.org/abs/1703.05175