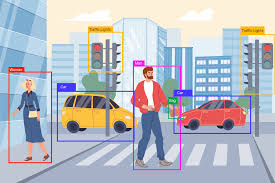
**Báo cáo AI/Machine Learning tuần 5**

1. **Bài toán object detection và image segmentation.**

* Object detection là bài toán xác định và phân loại các vật thể xuất hiện có trong một bức ảnh.
* Thuật toán object detection có đầu vào là 1 ảnh, đầu ra là các hộp bao quanh (bounding box) xung quanh mỗi đối tượng trong ảnh, đồng thời gán cho nó một nhãn (label) tương ứng với loại đối tượng đó.



* Ứng dụng của bài toán object detection:
* **Xe tự lái:** Phát hiện các phương tiện khác, người đi bộ, biển báo giao thông,.. qua đó đưa ra xử lý tình huống giao thông ngay lập tức.
* **Giám sát an ninh:** Phát hiện người lạ xâm nhập, đối tượng khả nghi từ đó đưa ra cảnh báo.
* **Tìm kiếm hình ảnh:** Tìm kiếm các hình ảnh chứa đối tượng cụ thể.
* Image Segmentation (phân đoạn vật thể trong ảnh) là một nhiệm vụ phức tạp hơn, ở đây là chia hình ảnh thành các vùng đồng nhất theo những tiêu chí nhất định như màu sắc, kết cấu, cùng đối tượng, mục tiêu là tạo ra 1 bản đồ phân đoạn, trong đó mỗi pixel được gán nhãn tương ứng với vùng mà nó thuộc về.

A person with her tongue out

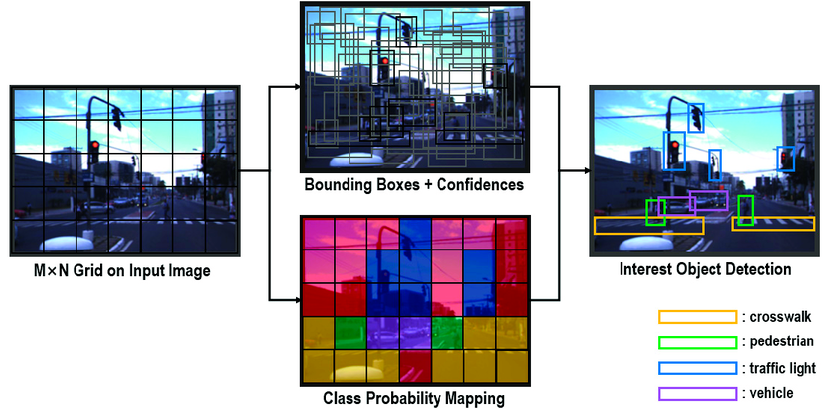
Description automatically generated

* Ứng dụng:
* Xử lý ảnh y tế: Phân đoạn các cơ quan trong cơ thể từ hình ảnh X-quang, MRI.
* Chỉnh sửa ảnh phức tạp: Tách đối tượng ra khỏi nền, lựa chọn đối tượng từ ảnh.
* So sánh Object Detection và Image Segmentation.

|  |  |
| --- | --- |
| Object Detection | Image Segmentation |
| Xác định đối tượng, phân loại vị trí của đối tượng. | Chia ảnh thành các vùng đồng nhất |
| Bounding box và labels của từng vật thể xuất hiện trong ảnh | Bản đồ phân đoạn bao quanh từng vật thể trong ảnh |
| Độ phức tạp thấp hơn | Độ phức tạp cao hơn |

1. **Model thường sử dụng và phân tích.**

* Ta có nhiều thuật toán để giải quyết 2 bài toán kể trên, tuy nhiên ta sẽ đề cập đến mô hình thông dụng có thể giải quyết được cả 2 bài toán kể trên là YOLO.
* YOLO (You Only Look Once) là một thuật toán phát hiện đối tượng nổi tiếng, được thiết kế để thực hiện phát hiện đối tượng trong thời gian thực. Tuy nhiên, YOLO ban đầu chủ yếu tập trung vào bài toán object detection, nghĩa là xác định vị trí và phân loại các đối tượng trong một hình ảnh bằng cách vẽ các bounding box xung quanh chúng.
* Cách YOLO giải quyết bài toán nhận diện đồ vật:
* Ảnh đầu vào được truyền qua 1 CNN để trích xuất các đặc trưng thuộc tính có trong ảnh.
* Các đặc trưng có được sẽ được truyền qua 1 lớp fully-connected, dự đoán thuộc tính đó thuộc về lớp đối tượng nào và tọa độ của bounding box tương ứng với đối tượng đó.
* Chia để trị: Chia ảnh ra thành lưới các ô có diện tích bằng nhau.
* Trong mỗi ô, hệ thống dự đoán xác suất mà mỗi ô đó thuộc về bounding box của đối tượng nào, từ đó cho ra xác suất mà ô thuộc đối tượng nào.
* Đầu ra của hệ thống là tập hợp các bounding boxes và xác suất thuộc về từng lớp của mỗi ô.
* Các bounding boxes sau đó được lọc thông qua thuật toán hậu xử lý non-max suppression để chọn ra box với xác suất cao nhất
* Đầu ra cuối cùng là tập hợp dự đoán bằng các bounding boxes, và nhãn các lớp cùng xác suất cho từng đối tượng xuất hiện trên ảnh.



* YOLO ban đầu không được thiết kế để thực hiện image segmentation, tức là chia một hình ảnh thành các vùng đồng nhất tương ứng với các đối tượng khác nhau. Tuy nhiên, với một số điều chỉnh và kết hợp với các kỹ thuật khác, YOLO có thể được sử dụng để giải quyết bài toán này. Phiên bản mới nhất của YOLO, YOLOv8, đã giới thiệu một mô hình chuyên biệt cho bài toán image segmentation. Nó thực hiện cả ba nhiệm vụ trong một lần dự đoán duy nhất:
* **Phát hiện đối tượng:** Xác định vị trí của các đối tượng bằng cách tạo ra các bounding box.
* **Phân loại đối tượng:** Xác định nhãn của từng đối tượng được phát hiện.
* **Phân đoạn đối tượng:** Tạo ra các mask (mặt nạ) để xác định chính xác hình dạng của từng đối tượng.
* Cách YOLO hoạt động trong bài toán image segmentation:
* Hầu hết các bước trong quá trình YOLO giải quyết bài toán phân đoạn hình ảnh cũng tương tự như khi giải quyết bài toán nhận diện vật thể.
* Điểm khác biệt ở chỗ mặt nạ phân đoạn (segmentation mask prediction): Ngoài việc dự đoán hộp bao quanh, YOLO còn dự đoán một mặt nạ phân đoạn (segmentation mask) cho mỗi vật thể. Mặt nạ này giống như một bản đồ nhiệt, cho biết khả năng mỗi pixel trong hình ảnh thuộc về vật thể đó.
* Hộp bao quanh cunc cấp vị trí xấp xỉ vật thể, mặt nạ phân đoạn cung cấp thông tin chi tiết về hình dạng và ranh giới của vật thể
* Bằng cách kết hợp cả hai, YOLO có thể thực hiện phân đoạn cấp độ instance, tức là xác định và phân cách từng vật thể riêng lẻ trong hình ảnh.

1. **Object detection với Faster CNN**

* Với bài toán object detection, nếu chỉ sử dụng CNN thì sẽ không giải quyết được do không thể xây dựng mô hình mà không biết có bao nhiêu đối tượng trong ảnh.

🡪 Cần một mô hình hiệu quả hơn, regional CNN ra đời.

* Mô hình RCNN có cơ chế hoạt động dựa trên thuật toán selective search algorithm, cụ thể:
* Dùng Selective Search algorithm để lấy ra khoảng 2000 bounding box trong input mà có khả năng chứa đối tượng. Về thuật toán này, đầu tiên ảnh được segment qua thuật toán Graph Based Image Segmentation, từ đó nhóm các vùng màu với nhau để làm region proposal 🡪 quay về bài toán phân loại các region proposal.

A palm tree on a beach

Description automatically generated

* Mỗi region proposal được đưa vào một mạng CNN đã được huấn luyện trước để trích xuất các đặc trưng sâu. Các đặc trưng trích xuất được đưa vào một bộ phân loại (ví dụ như SVM) để xác định xem region proposal đó có chứa vật thể hay không và nếu có thì đó là vật thể thuộc lớp nào.
* Điều chỉnh Bounding Box (optional): Nếu region proposal được phân loại là một vật thể, một bộ hồi quy sẽ được sử dụng để điều chỉnh kích thước và vị trí của bounding box sao cho phù hợp hơn với vật thể.

A diagram of a person in a cowboy hat

Description automatically generated

* Ưu điểm chính: Giải quyết được bài toán object detection với độ chính xác cao.
* Tuy nhiên, 1 số nhược điểm tồn tại như quá trình trích xuất đặc trưng cho từng region proposal là rất tốn kém về mặt tính toán, dẫn đến tốc độ xử lý chậm. Hay RCNN bao gồm nhiều giai đoạn xử lý riêng biệt, làm tăng độ phức tạp của mô hình.
* Để khắc phục những hạn chế của RCNN, các nhà nghiên cứu đã tiếp tục phát triển các mô hình cải tiến như Fast R-CNN, Faster R-CNN,...

1. **Cách train model và các thông số đánh giá độ hiệu quả của 1 model.**

* Quá trình huấn luyện mô hình để giải quyết bài toán object detection thường sẽ trải qua quá trình sau đây:
* Cần một bộ dữ liệu lớn, đa dạng, bao gồm nhiều hình ảnh với các vật thể khác nhau ở nhiều góc độ, kích thước và điều kiện ánh sáng khác nhau. Chia bộ dữ liệu thành 3 tập: train, validation và test.
* Mỗi hình ảnh trong tập train cần được đánh dấu các bounding box (khung bao) và nhãn tương ứng cho từng vật thể. 1 số ứng dụng phổ biến như LabelImg hoặc roboflow,....
* Xây dựng kiến trúc mô hình dựa trên yêu cầu bài toán và tài nguyên có sẵn.
* Huấn luyện: chuyển dữ liệu đã gán nhãn sang định dạng phù hợp với framework đã chọn (tensorflow, pytorch,keras) và truyền vào mô hình.
* Định nghĩa hàm mất mát để đánh giá sự khác biệt giữa nhãn, bounding box dự đoán và thực tế.
* Sử dụng các thuật toán tối ưu hóa (gradient descent ,...) để cập nhật các tham số của mô hình.
* Tiếp tục huấn luyện và đánh giá mô hình đến khi kết thúc.
* Một số những thông số quan trọng trong việc đánh giá mô hình:
* **IoU (Intersection over Union):** Đo lường mức độ trùng khớp diện tích giữa bounding box dự đoán và bounding box ground truth. Chỉ số IoU trong khoảng [0,1], càng gần 1 thì bounding box dự đoán càng gần ground truth.
* **mAP (mean Average Precision):** Đo lường hiệu năng tổng thể của mô hình trên tất cả các lớp.
* **Precision:** Tỷ lệ các dự đoán đúng trong tổng số dự đoán. 🡪 đo độ chính xác các dự đoán, precision cao có thể dự đoán ít mà đúng nhiều nhưng có khả năng bỏ qua các dự đoán vật thể bé, khuất tầm.
* **Recall:** Tỷ lệ các vật thể thực sự được phát hiện đúng trong tổng số vật thể thực sự. 🡪 Đo khả năng phát hiện đầy đủ các đối tượng. Recall cao chứng tỏ mô hình phát hiện được hầu hết các đối tượng, nhưng cũng có thể đưa ra nhiều dự đoán sai.

1. **Background Subtraction.**

* Background Subtraction là quá trình tách đối tượng chuyển động khỏi nền tĩnh trong một hình ảnh hoặc video. Nói cách khác, chúng ta muốn tìm ra những điểm ảnh thay đổi so với hình ảnh nền để xác định các đối tượng đang di chuyển.
* Mục tiêu khi thực hiện background substraction là năhmf xác định, nhận biết và phân tách các đối tượng đang chuyển động trên nền tĩnh, từ đó hướng đến các tác vụ cao hơn: theo dõi chuyển động đối tượng trong 1 thời gian hoặc phân tích hành vi, xu hướng của vật thể.
* Một trong những kĩ thuật phổ biến nhất để thực hiện quá trình background substraction là frame differencing ( vi phân hình ảnh) với cách hoạt động như sau:
* Hệ thống lưu trữ 2 khung hình liên tiếp của video hoặc ảnh cần loại nền với ảnh nền tĩnh nếu sử dụng ảnh.
* Các giá trị pixel từ frame mới sẽ bị trừ đi tương ứng bới các pixel từ frame cũ, quá trình này cho ra 1 ảnh mới.
* Các giá trị pixel của ảnh đó sẽ được so sánh với một ngưỡng cố định, nếu lớn hơn chứng tỏ pixel đó thuộc về đối tượng di chuyển/ không phải nền.

A diagram of a flowchart

Description automatically generated

* Ưu điểm: Đơn giản, dễ triển khai và tính toán hiệu quả, thích hợp cho các hệ thống theo dõi chi phí thấp cần phát hiện đối tượng chuyển động đột ngột.
* Nhược điểm của thuật toán này là thường nhạy cảm với nhiễu dẫn đến phát hiện giả. Ngoài ra nó khó phát hiện các đối tượng di chuyển chậm hoặc có sự thay đổi nhỏ về mặt di chuyển, dễ nhầm lẫn đấy là thành phần background.