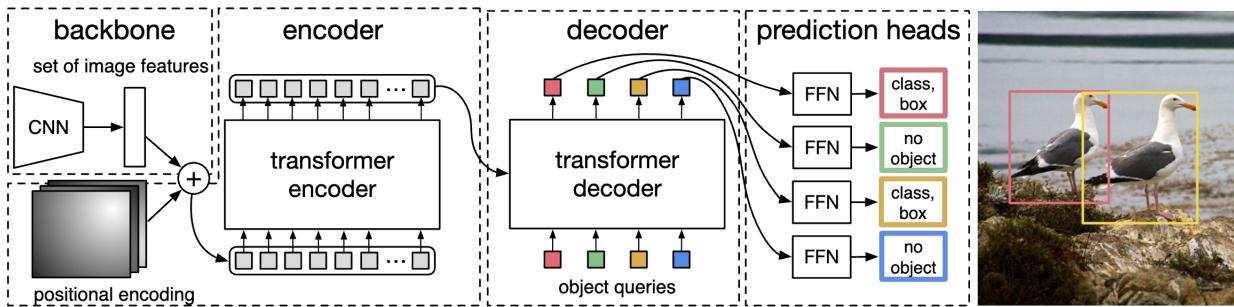


# Exercise Object Detection - Description

Ngày 1 tháng 1 năm 2024

## 1 Giới thiệu Detection Transformer (DERT)

Giữa năm 2020, bài báo về Detection Transformer (DERT) xuất hiện tạo ra một phương pháp tiếp cận mới trong bài toán Object Detection, tận dụng sức mạnh từ transformer.



Hình 1: Kiến trúc DERT.

### 1.1 Giới thiệu

Transformer đã tạo ra sự đột phá trong lĩnh vực Natural Language Processing (Transformer Encoder-Decoder, BERT, ...) và gần đây đã đạt được những bước tiến quan trọng trong Computer Vision (Vision Transformer). Detection Transformer (DERT) là một biểu hiện của xu hướng này, áp dụng kiến trúc transformer vào bài toán Object Detection.

### 1.2 Kến trúc DERT

Kiến trúc DERT có thể được chia thành ba phần: backbone (CNN), transformer encoder, và transformer decoder. Backbone là các model CNN (ResNet-50, ResNet-101) dùng để trích xuất feature map từ hình ảnh input ban đầu.

Các feature map sau đó được flatten và cộng thêm positional encoding rồi đưa vào transformer encoder để tổng hợp thông tin từ các phần khác nhau trong ảnh. Transformer decoder nhận vào input và output của encoder và các object query để tổng hợp và output ra các bounding box. Các object query đóng vai trò như những người đánh giá khác nhau, ví dụ object query xác định các objet ở vị trí góc bên trái của bức ảnh thì các object query khác sẽ tập trung vào các phần khác của bức ảnh. Cuối cùng, output của decoder sẽ được đưa vào head để predict ra class object và vị trí bounding box.

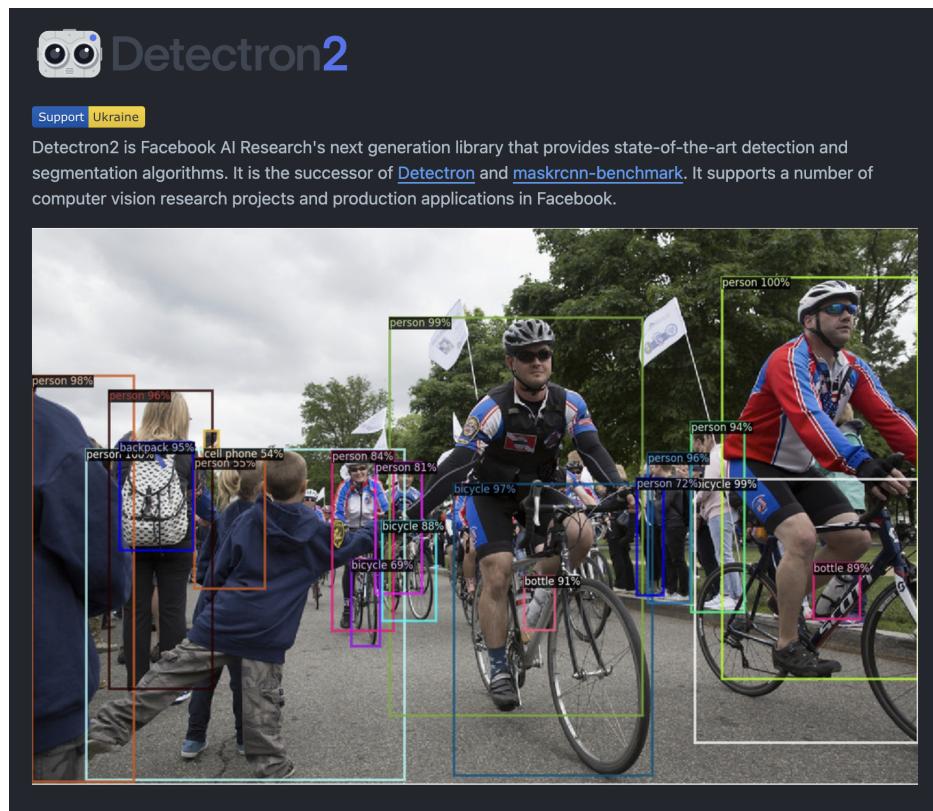
### 1.3 Ưu điểm của DERT

DERT mang lại nhiều ưu điểm so với các phương pháp object detection khác cùng thời điểm đó. Bao gồm khả năng xử lý hiệu quả hơn với các đối tượng có kích thước đa dạng, cải thiện hiệu suất trong các cảnh phức tạp, và tiềm năng tổng quát hóa tốt hơn nhờ cơ chế dựa trên attention trong transformer. DERT cũng không cần sử dụng các thành phần hand-designed vốn tồn tại trong các model khác (NMS, anchor box, ...).

## 2 Một số library training Object Detection model

### 2.1 facebookresearch/detectron2

[Detectron2](#) là một github repo của Facebook AI Research (FAIR), cung cấp các SoTA model về Object detection và Segmentation.



### 2.2 PaddlePaddle/PaddleDetection

[PaddleDetection](#) là github repo của PaddlePaddle, cung cấp các pipeline end-to-end object detection, segmentation, tracking, ... với hơn 30 model, algorithm và hơn 300 pre-trained model. Đặc biệt, PaddleDetection cung cấp các mô hình SOTA công nghiệp high-performance & light-weight trên server và thiết bị mobile. PaddleDetection cung cấp nhiều phương pháp data augmentation, config các network, loss function. PaddlePaddle cũng cung cấp nhiều trường hợp và tutorial phong phú để tăng tốc ứng dụng của model.



## 2.3 Ultralytics

Ultralytics là source gốc của phiên bản YOLO mới nhất (YOLOv8) và các giải pháp cho các bài toán khác: segmentation, pose estimation, và tracking. Bên cạnh đó, ultralytics còn cung cấp các giải pháp cho model training, prediction, và deployment.

