YOLO

You Only Look Once (YOLO) is a new approach to object detection. The existing approach to object detection usually repurposes classifiers to perform detection. However, it struggles to perform complex tasks. YOLO reframes the object detection as a single regression problem, straight from image pixels to bounding box coordinates and class probabilities, extremely fast and accurate. It can process streaming video in real-time with less than 25 milliseconds of latency. YOLO is trained on a loss function that directly corresponds to detection performance and the entire model is trained jointly. It sees the entire image during training and test time so it implicitly encodes contextual information about classes as well as their appearance, just like humans glancing at an image and instantly knowing what objects are in the image, where they are, and how they interact. YOLO will resize the input image to 448*448 and divide it into an S × S grid and for each grid cell predicts B bounding boxes, confidence for those boxes, and C class probabilities. YOLO runs a single convolutional network on the image, and thresholds the resulting detections by the model’s confidence. YOLO's error rate in false detection (detecting the background as an object) can be reduced by more than half in training. While it can quickly identify objects in images it struggles to precisely localize some objects that lags behind state-of-the-art detection systems, especially small ones.