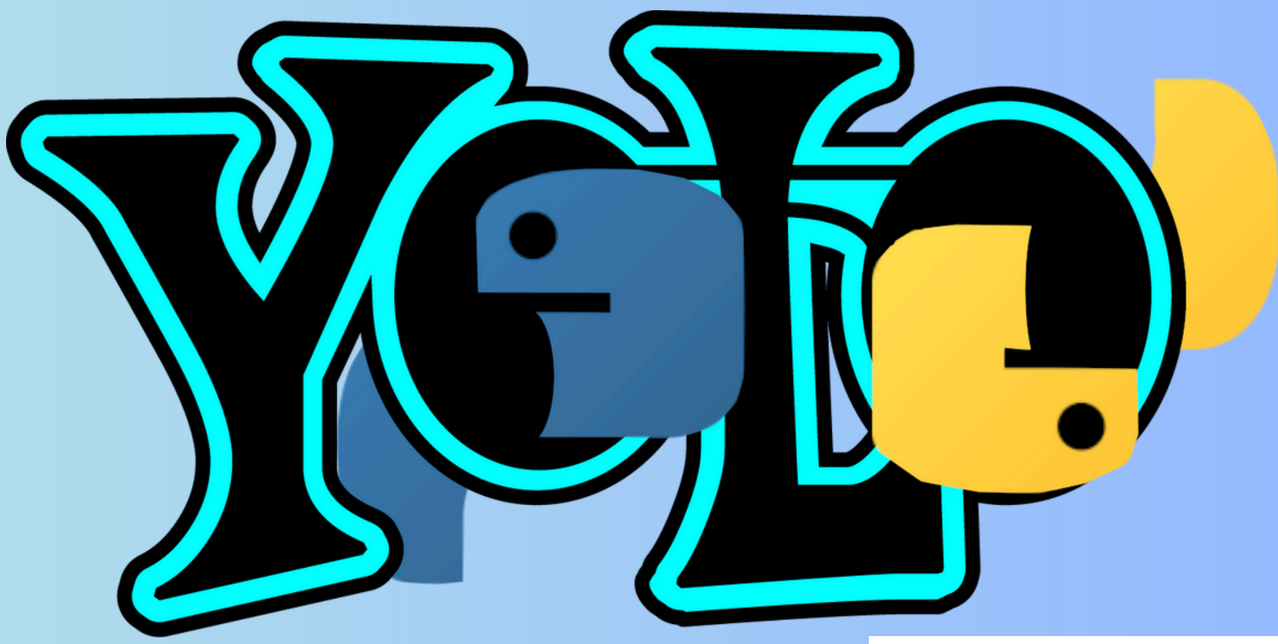
Mediapipe analogs

YOLO, OpenPose, DeepLabCut, Dlib, AlphaPose

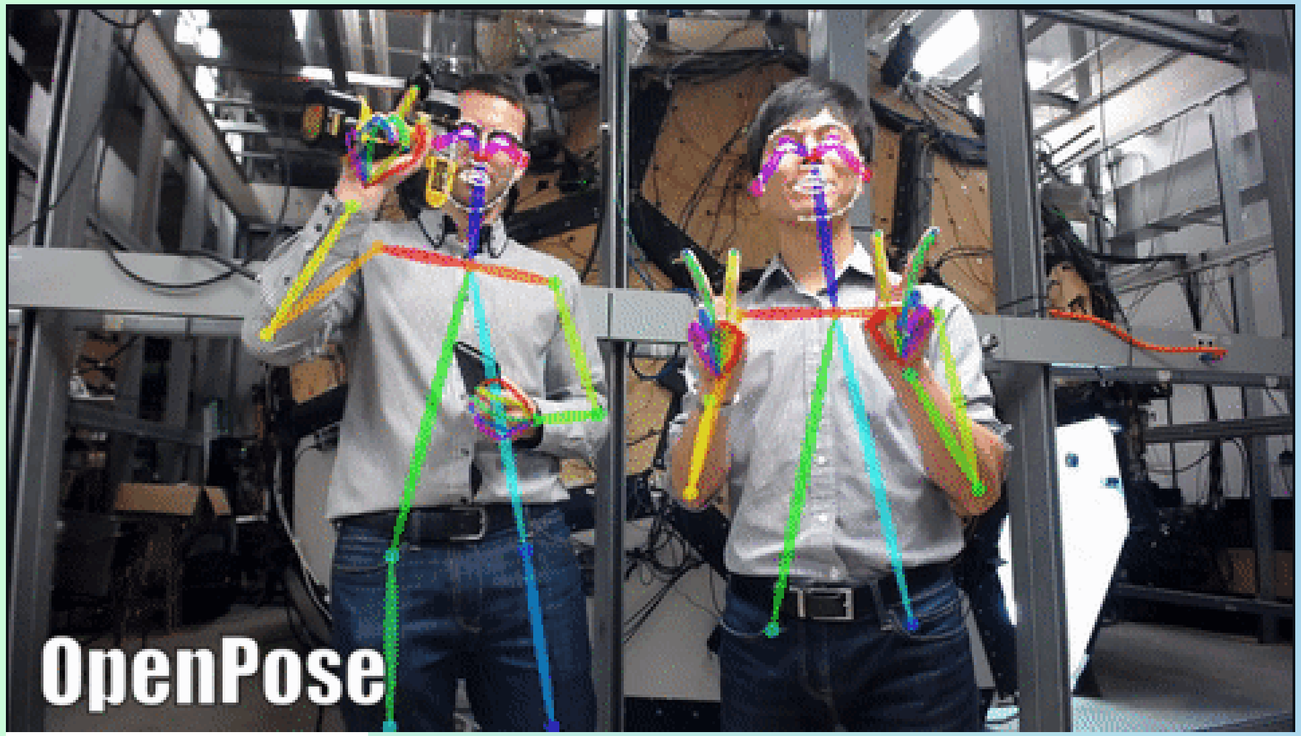
Presentaed by Zulkhairatuly Yerlan

YOLO (You Only Look Once)

It is an object detection system developed by Joseph Redmon in 2015. It is designed for real-time object detection, using a single-pass deep learning model. It supports multiple versions, from YOLOv1 to YOLOv8. It is commonly used in autonomous vehicles, security surveillance, and retail checkout systems. The most recent version is YOLOv8, released in 2023.



OpenPose

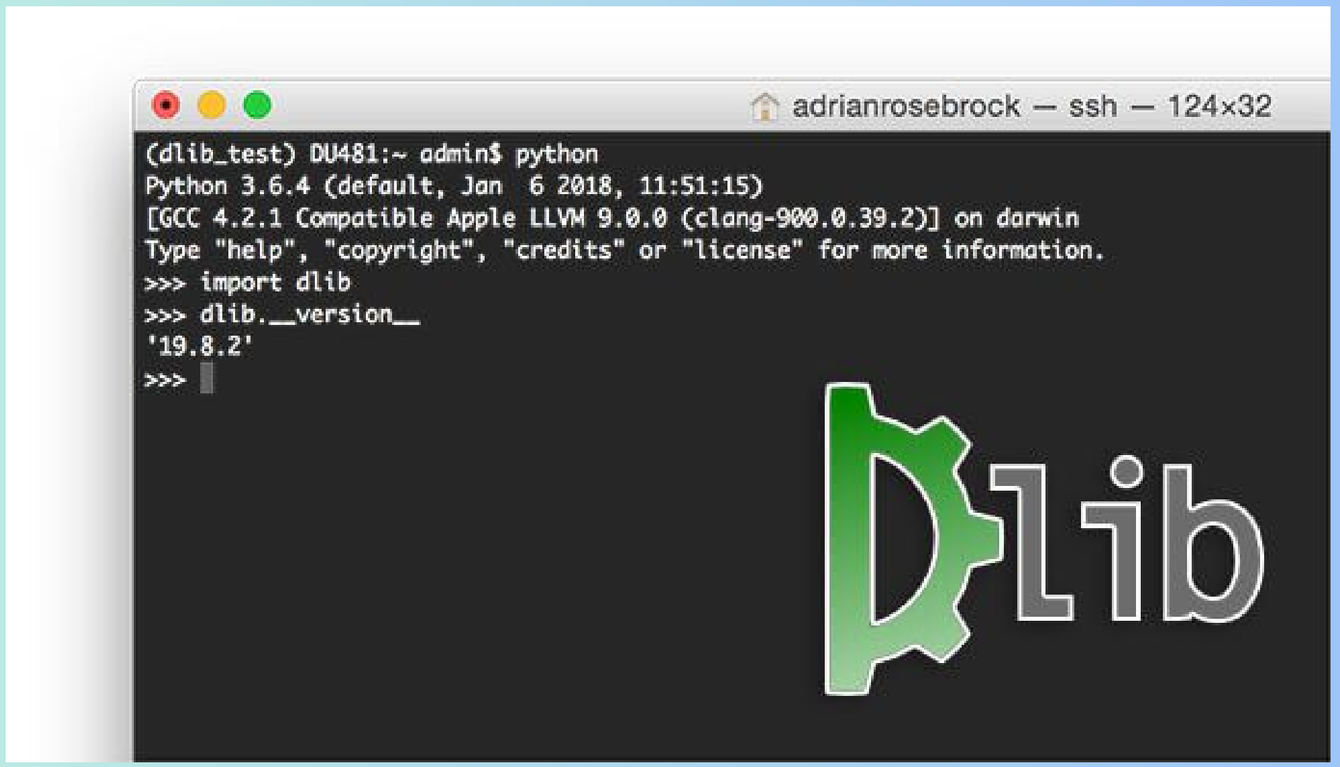


It is a pose estimation method developed by Carnegie Mellon University. It detects full-body skeletons for multiple people in an image and works on both humans and animals. It uses a bottom-up approach for keypoint detection. Applications include sports analytics, fitness tracking, and gesture-based interaction. It is known for accurately tracking multiple people.

DeepLabCut

It is a pose estimation framework developed by Mackenzie Mathis Lab, designed for tracking both animals and humans. It uses deep learning to detect keypoints and requires minimal training data. It is used in behavioral neuroscience, animal movement studies, and biomechanics research. It is highly accurate for tracking non-human subjects.

Dlib



It is a face detection and landmark tracking library developed by Davis E. King. It provides 68 facial landmark detection, pre-trained models, and efficient CPU-based processing. It is used in facial recognition, emotion analysis, and eye-tracking applications. It is lightweight and works well on low-power devices.

AlphaPose

It is a multi-person pose estimation system developed by MVIG-SYSU Lab. It provides high accuracy in tracking complex human poses and can detect occluded body parts. It is considered more effective than OpenPose in dense crowds. It is used in sports biomechanics, dance movement analysis, and AI-assisted rehabilitation.

Why YOLO

Real-Time Speed

YOLO's architecture allows it to process images much faster than traditional object detection models. This speed is crucial for applications requiring real-time analysis, such as autonomous driving and video surveillance.

Single-Pass Detection

By evaluating the neural network only once per image, YOLO minimizes computational overhead. This efficiency is a significant advantage when processing high-resolution images or video streams.

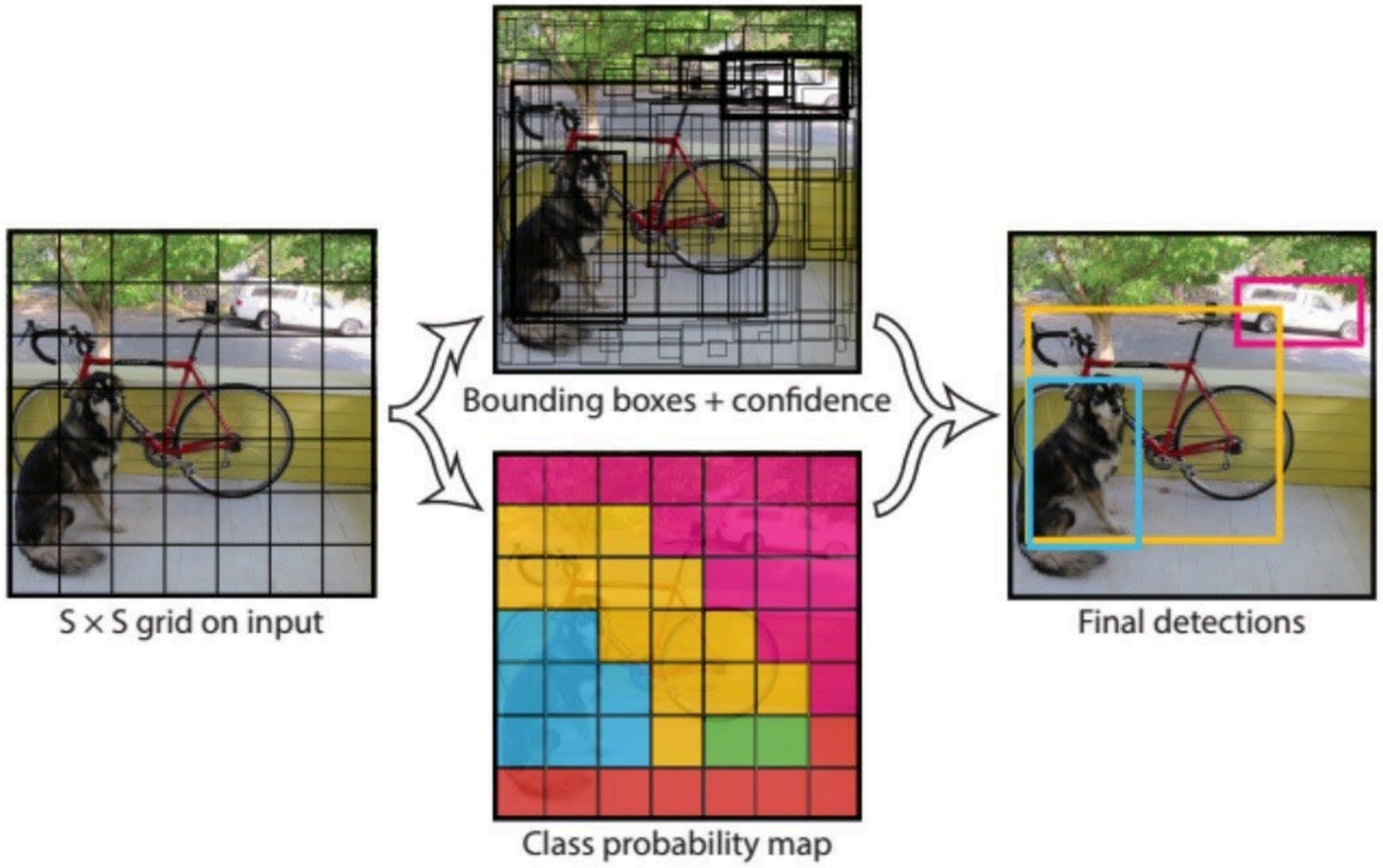


Figure 2: The Model. Our system models detection as a regression problem. It divides the image into an $S \times S$ grid and for each grid cell predicts B bounding boxes, confidence for those boxes, and C class probabilities. These predictions are encoded as an $S \times S \times (B * 5 + C)$ tensor.

Conclusihn

YOLO is a real-time object detection system known for its speed and efficiency. Unlike traditional methods, it processes an image in a single pass, making it suitable for applications requiring quick detection, such as autonomous vehicles, surveillance, and robotics. Over multiple versions, YOLO has improved in accuracy, efficiency, and functionality, supporting object detection, segmentation, and tracking. Despite its strengths, it may struggle with detecting small or overlapping objects. As research continues, newer versions are expected to enhance accuracy and performance, making YOLO a key technology in real-time computer vision.