Problem Understanding

The task was to implement a simplified version of an industrial scrap sorting vision pipeline. The system detects recyclable scrap items such as plastic, glass, and metal in real time, classifies them, and generates robotic pick-points. The solution was implemented using open-source tools and demonstrated on video or live camera feed.

Step-by-Step Breakdown

First, the dataset was prepared using the TACO dataset, which provides COCO-style bounding box annotations. A subset of around 150 images was selected across three categories: plastic, glass, and metal. Annotations were converted into YOLOv8 format with an 80/20 train-validation split. Next, a lightweight YOLOv8n model was fine-tuned on the dataset. Training was done for 10 epochs at an image size of 640x640. The trained model weights achieved reliable detection across the selected categories. A conveyor simulation was implemented using webcam or video feed. The trained YOLOv8 model was used to run real-time inference at around 5–10 FPS on CPU. Bounding boxes and confidence scores were drawn on detected objects. For each bounding box, the midpoint was calculated and visualized with a crosshair to simulate robotic pick-points. A simple dashboard overlay was added to show object counts per class and FPS, and latency was measured for each frame.

Key Decisions

The TACO dataset was chosen over TrashNet since it provides bounding box annotations. The YOLOv8n variant was selected for its lightweight architecture and fast inference speed. The pipeline was limited to three categories to simplify training while meeting assignment requirements.

Challenges and Learnings

The TACO dataset is large and has many fine-grained categories, so only a subset was selected for practical training. Real-time performance on CPU required using YOLOv8n and lowering the confidence threshold. This project provided experience in converting COCO annotations to YOLO format and integrating YOLOv8 with OpenCV.

Optimizations for Edge Deployment

Future optimization can include using YOLOv8n or YOLOv8s with quantization to reduce model size. The model can be exported to ONNX or TensorRT for deployment on Jetson Nano or Xavier. Video processing can be improved with multithreading, and the system could be integrated with ROS2 for robotic arm pick-and-place simulation.

Conclusion

The project successfully demonstrates a mini version of a scrap sorting vision pipeline. The trained YOLOv8 model detects recyclable materials in real time, generates pick-points, and overlays counts and performance metrics. With further optimization, the system can be deployed on low-power edge devices for industrial use.