Ssdlite mobilenet v2 Model Card

Model Card

Model Details

- The ssdlite_mobilenet_v2_INT8 model is an int8-quantized version of SSD Lite MobileNetV2, designed for fast and efficient object detection, particularly on resource-constrained devices. It is optimized for Intel's OpenVINO toolkit, leveraging quantization to reduce memory usage and improve inference speed while maintaining reasonable accuracy. It takes a 300x300 RGB image as input and outputs a set of detected objects, each represented by bounding box coordinates, a class label, and a confidence score.

Intended Use

- Our application uses this model for AI inferencing on input video and we collect metrics while the pipeline is running
- The model follows the SSD (Single Shot MultiBox Detector) architecture with a MobileNetV2 backbone, making it suitable for real-time applications such as surveillance, robotics, smart cameras, and IoT edge devices.

Training and validation data

- We are not training or validating this model in our reference implementation

Ethical Considerations

- We are using person-bicycle-car-detection.mp4 from https://github.com/intel-iot-devkit/sample-videos as input video to test this application tool.
- We are not storing any person or user related personal information.

Caveats and Considerations

- The model's accuracy may vary depending on the quality and resolution of the input images. Ensure that the images used are of sufficient quality for reliable detection.
- Preprocess images to normalize lighting conditions and remove noise.

Quantitative Analysis

- We are not doing quantitative analysis in this application tool but we do display metrics mentioned below to the user.

Factors

- We are also not evaluating this model in this reference implementation

Metrics

- We are displaying metrics including throughout (FPS) and system level metrics: CPU/GPU utilization, memory utilization, CPU/GPU frequency, CPU/system temp, GPU power, GPU engine, and package power. In this application these metrics are collected and displayed to the user via gauges.