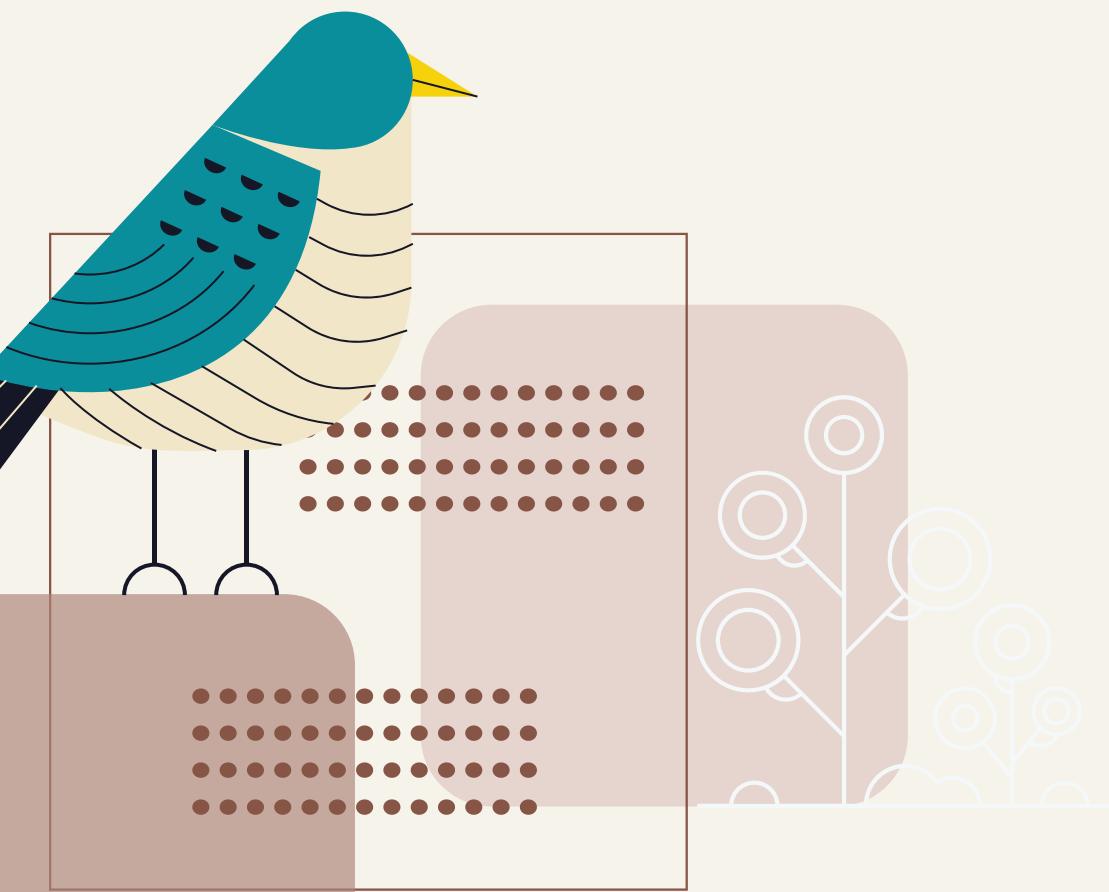


YOLO

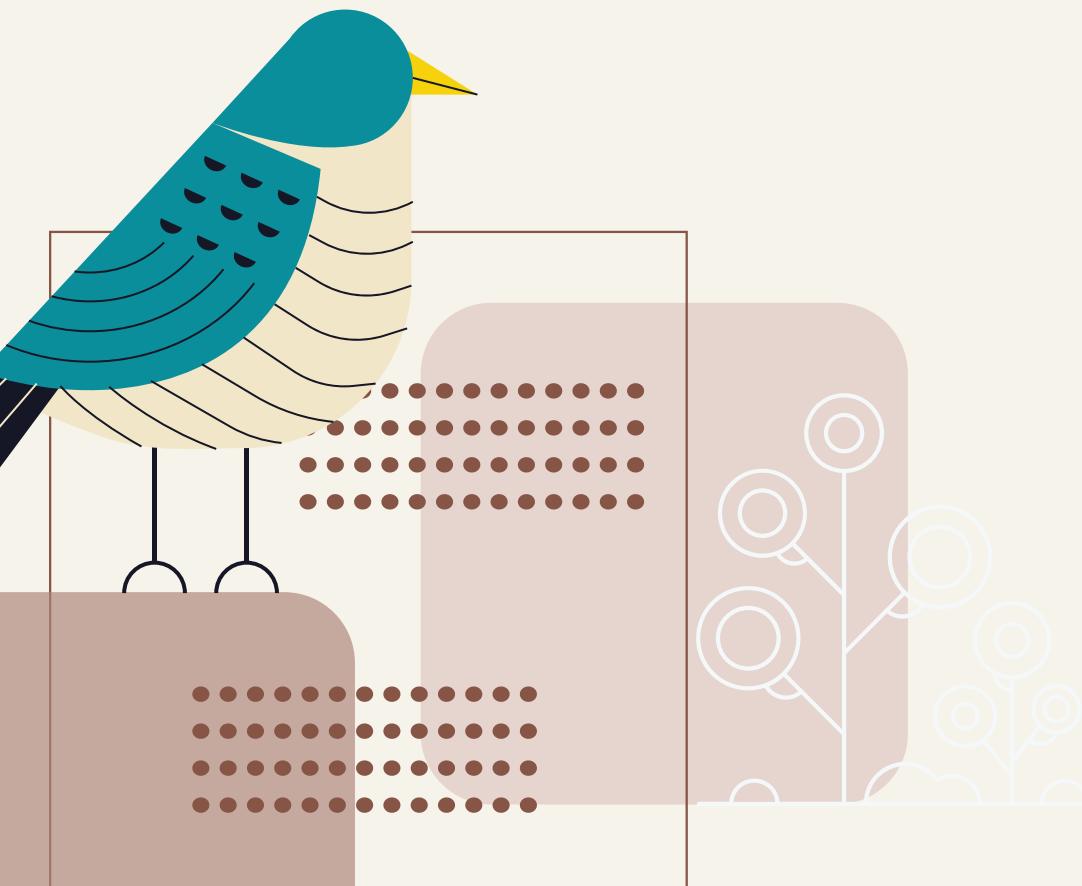
(You Only Look Once)

Presented By: Thanh Dang

QUANTIZATION MODEL



QUANTIZATION MODEL & AFTER CONVERSION



MODEL YOLOV5S.PT

Precision

0.671

Recall

0.519

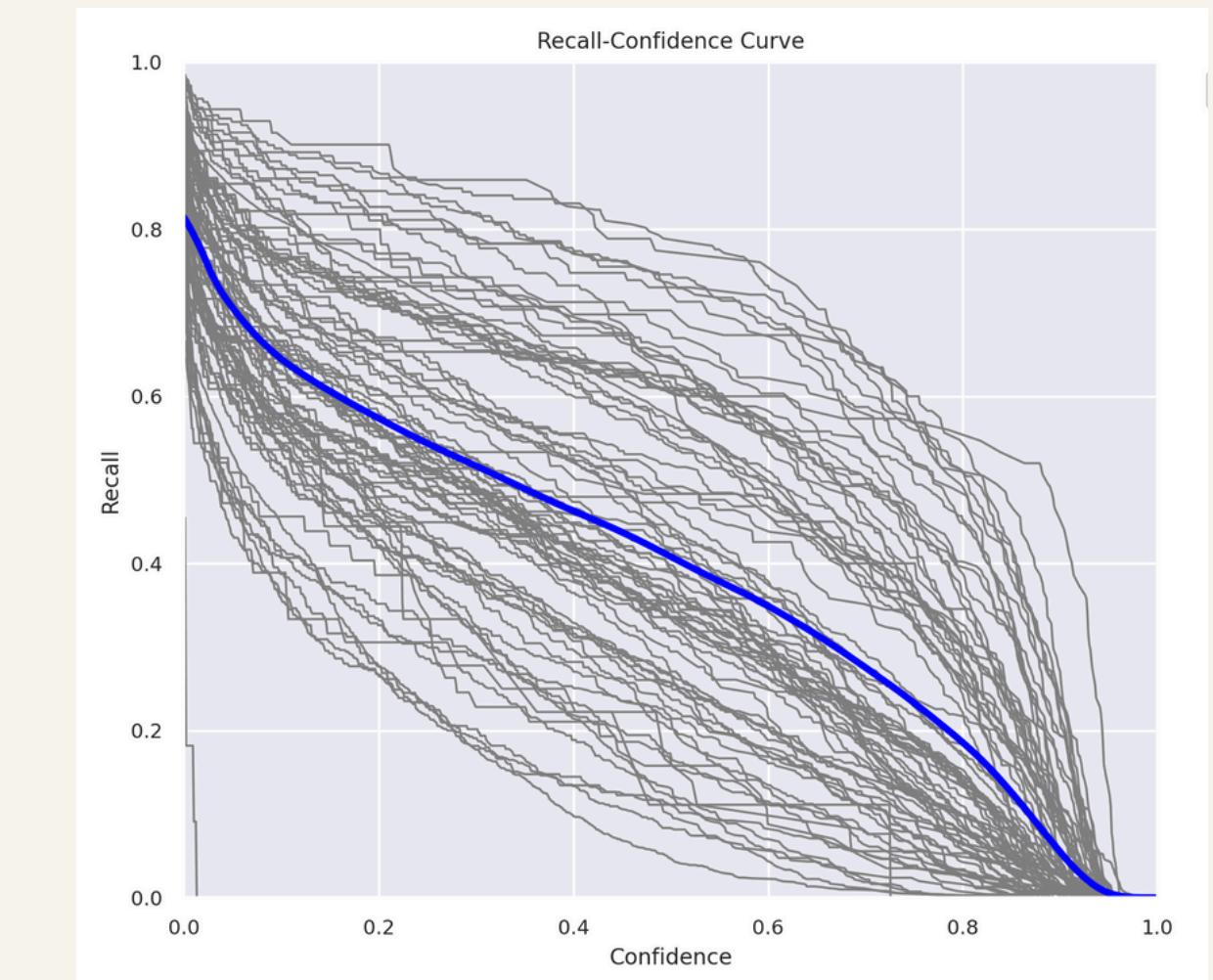
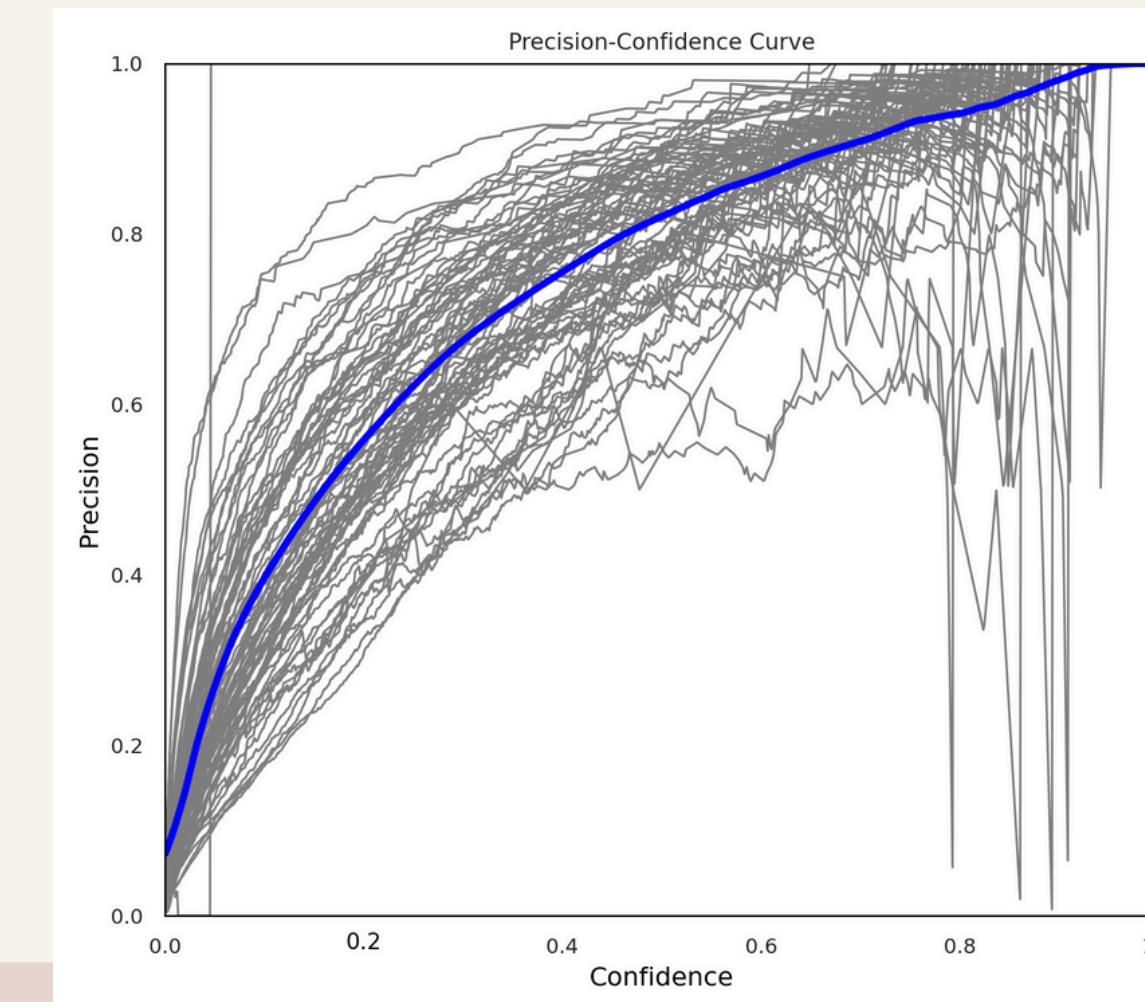
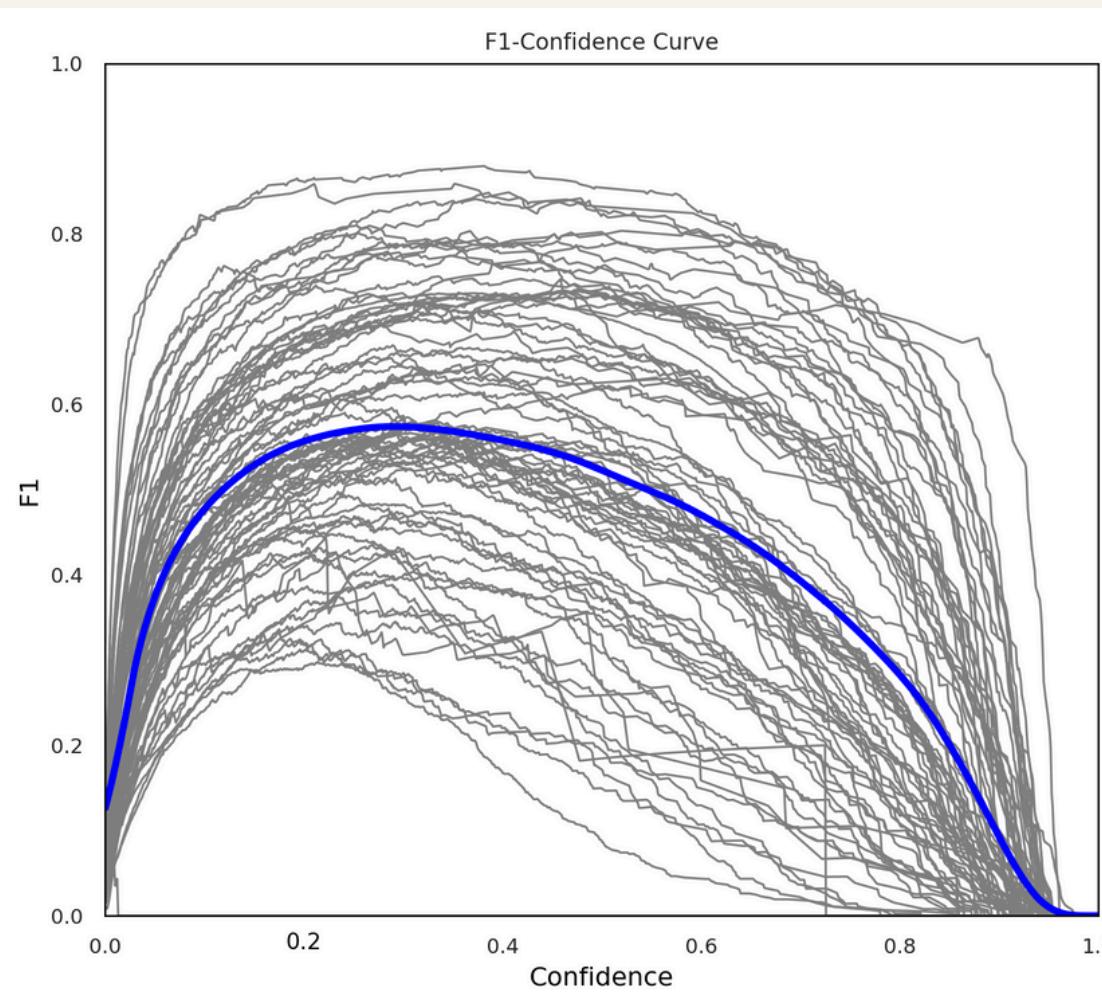
mAP50

0.566

mAP50-95

0.371

Speed: 0.1ms pre-process, 4.1ms inference, 2.6ms NMS per image



MODEL YOLOV5S.ONNX

Precision

0.669

Recall

0.661

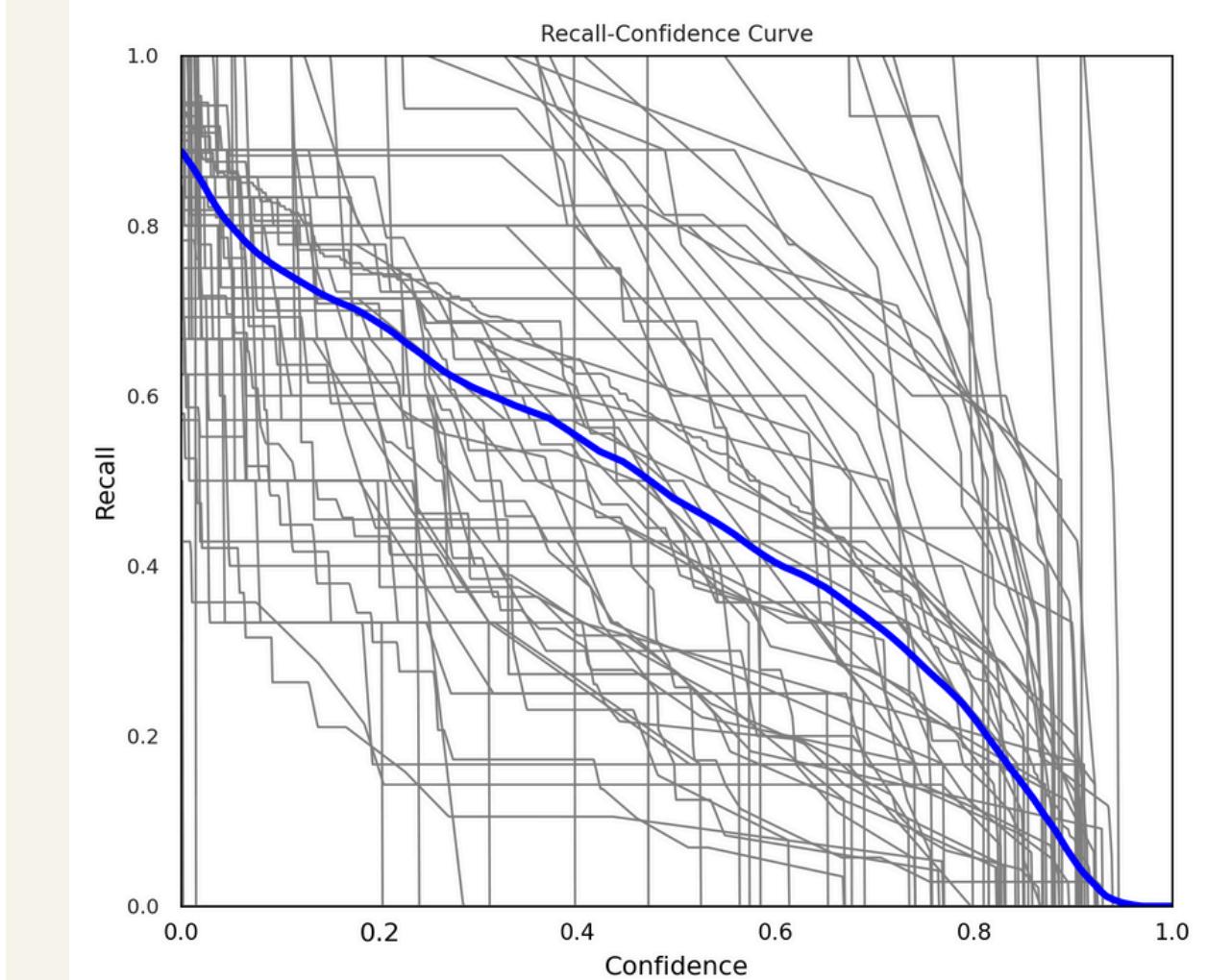
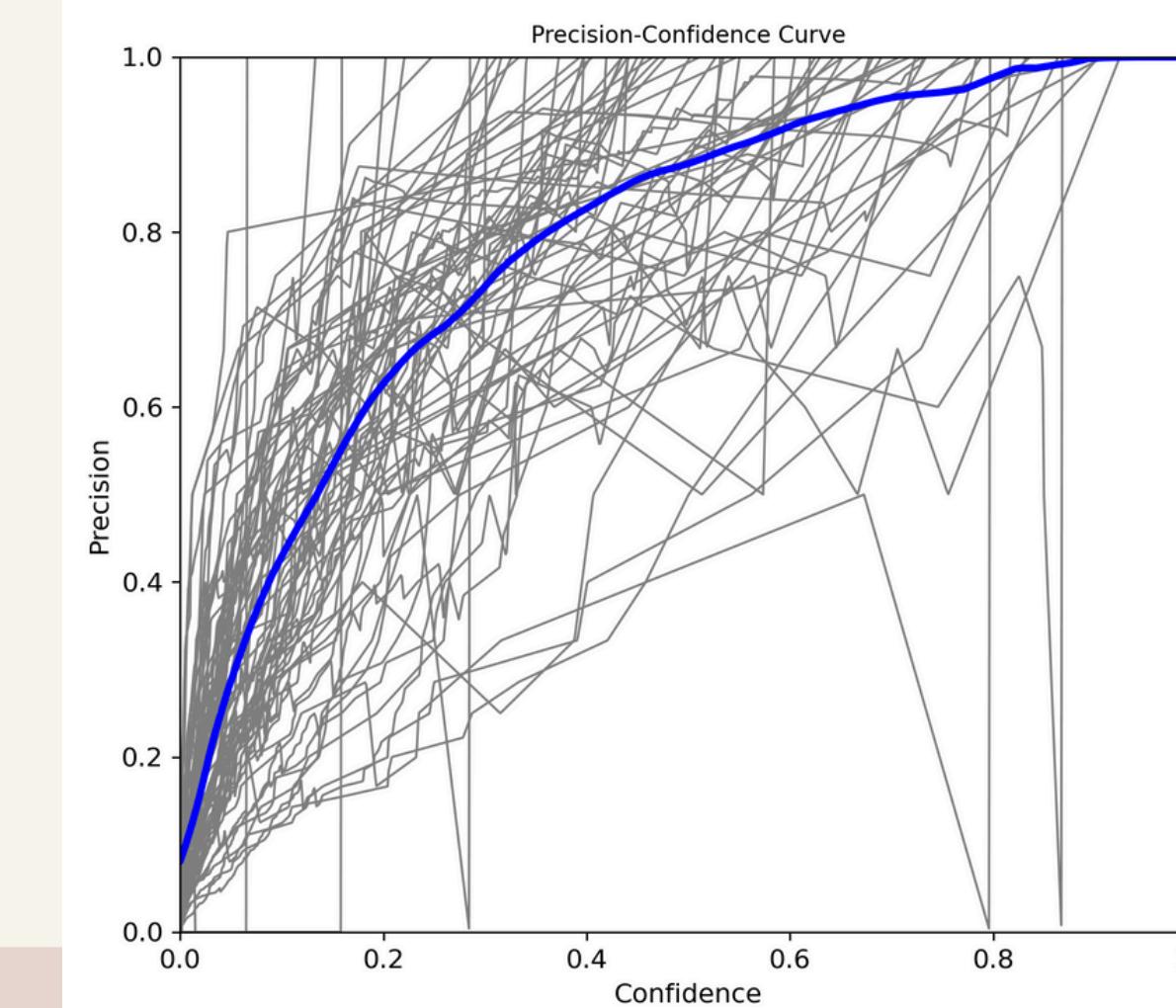
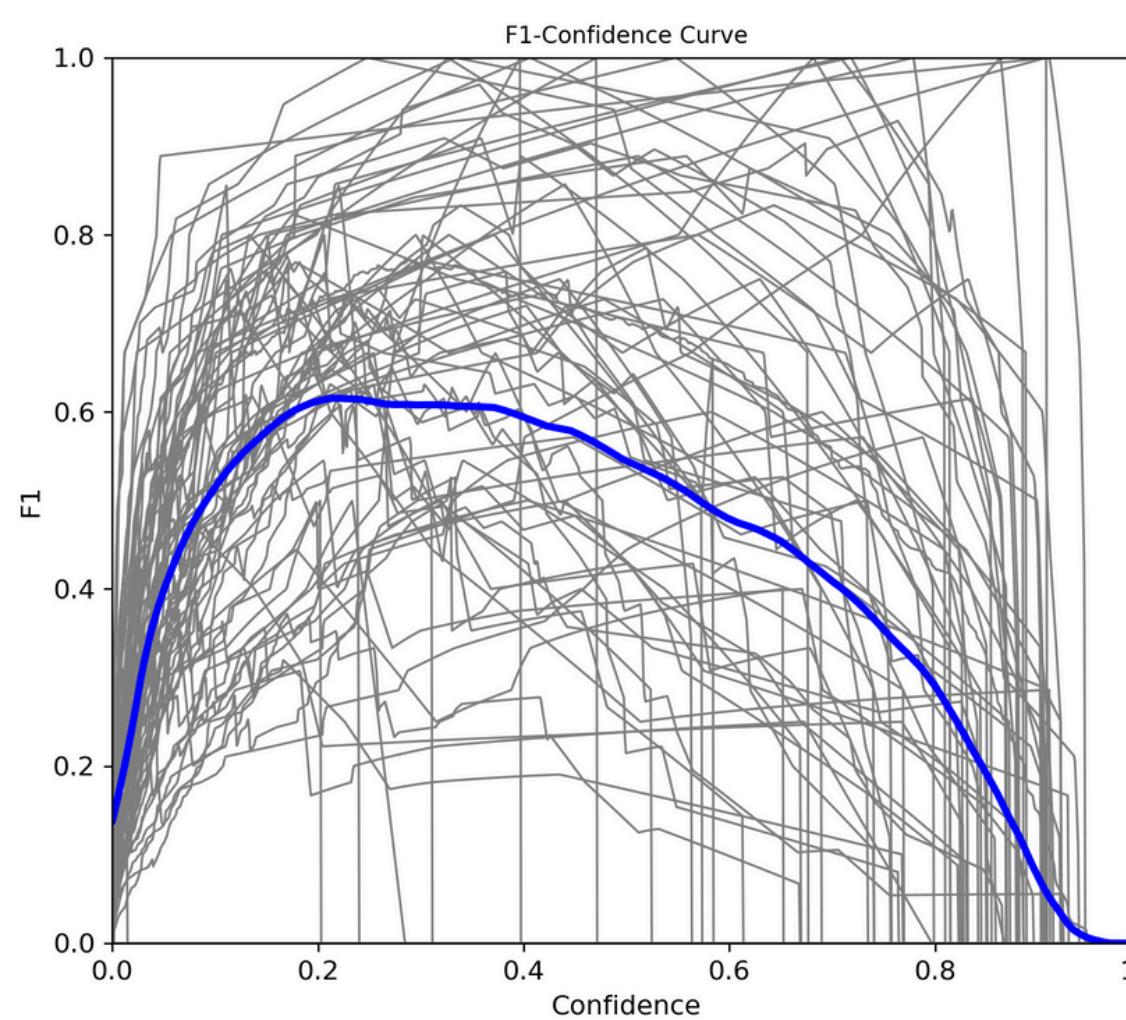
mAP50

0.712

mAP50-95

0.472

Speed: 0.4ms pre-process, 296.0ms inference, 7.8ms NMS/image



MODEL YOLOV5S.TORCHSCRIPT

Precision

0.669

Recall

0.661

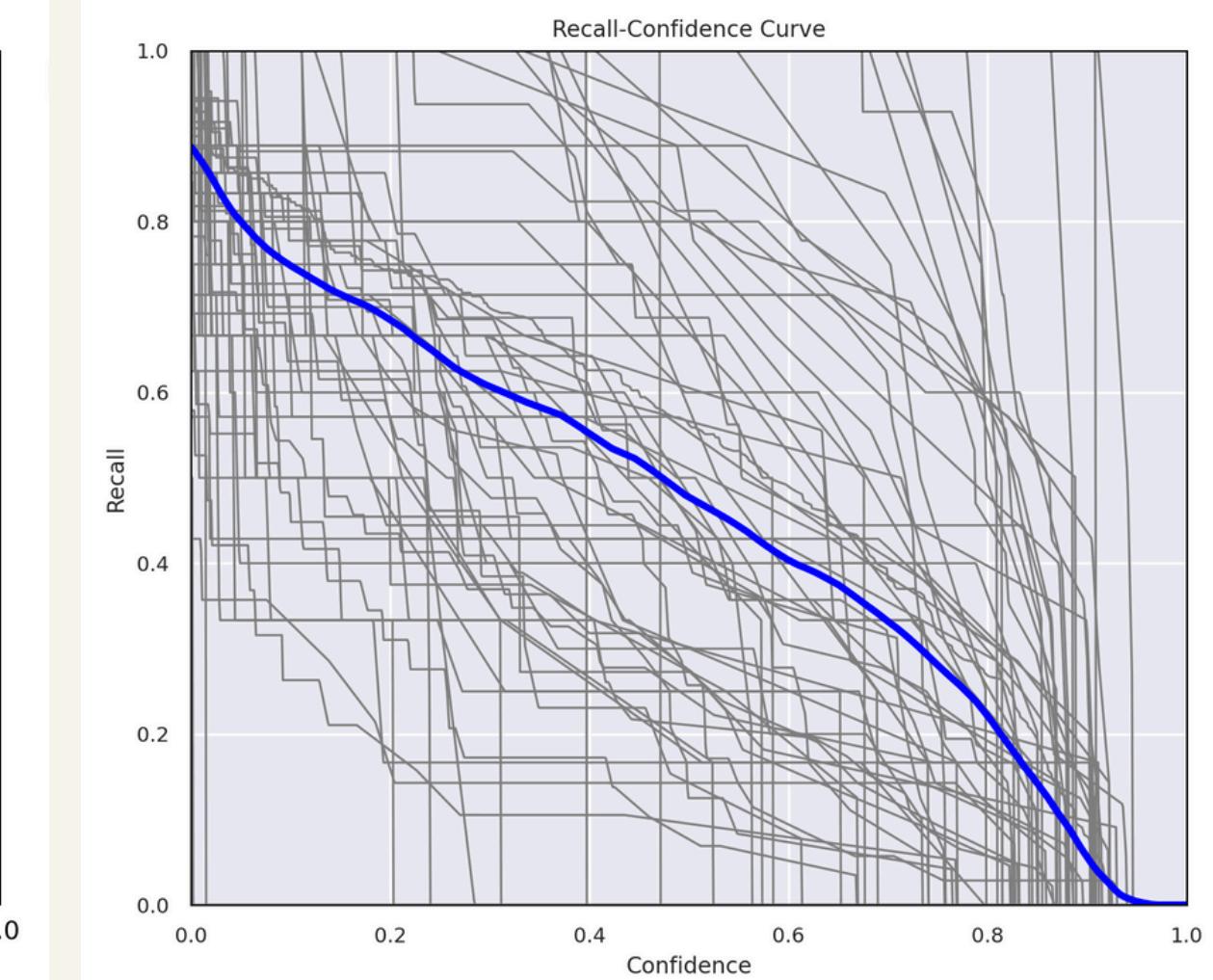
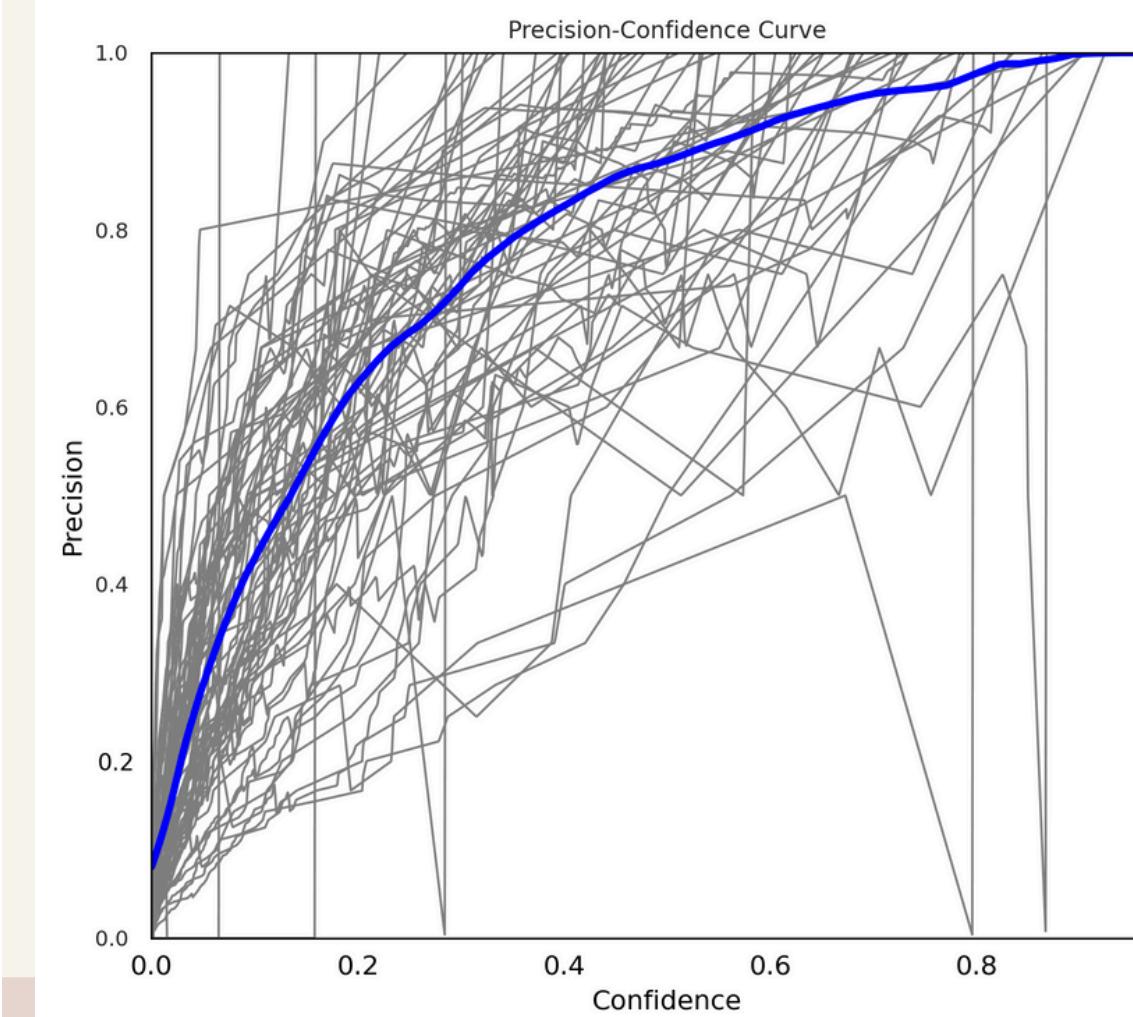
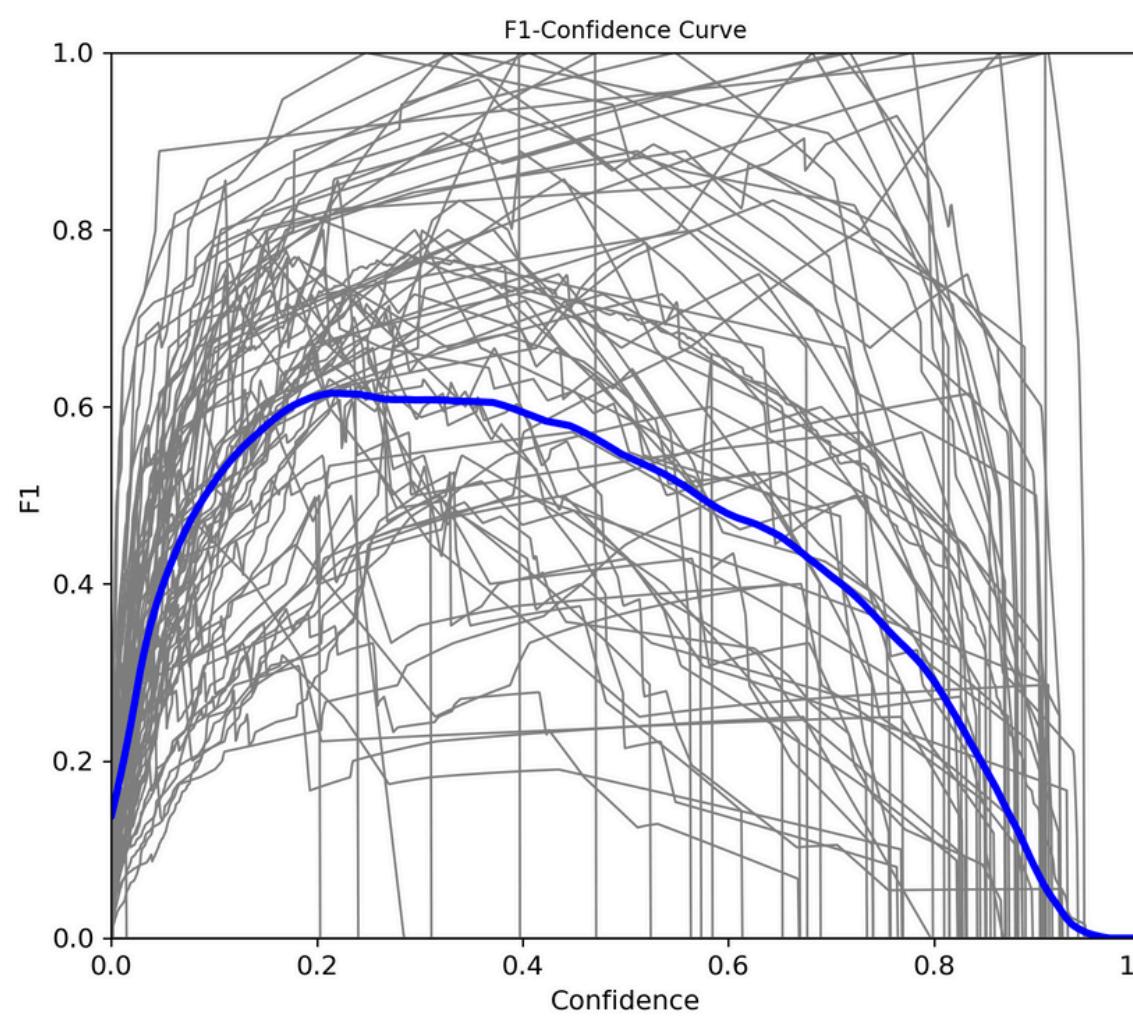
mAP50

0.712

mAP50-95

0.472

Speed: 0.5ms pre-process, 11.7ms inference, 10.0ms NMS per image



MODEL YOLOV5S-FP16.TFLITE

Precision

0.675

Recall

0.657

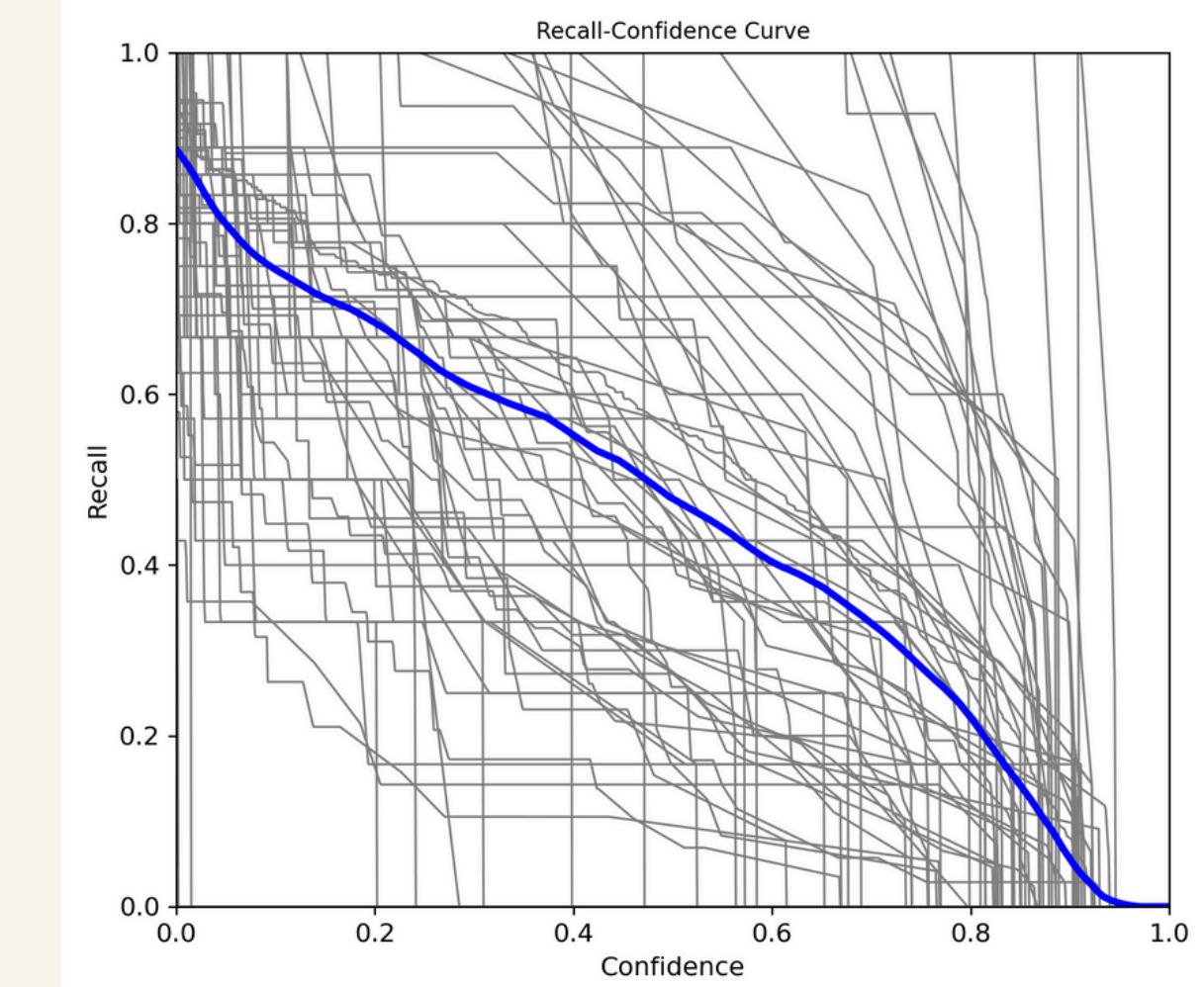
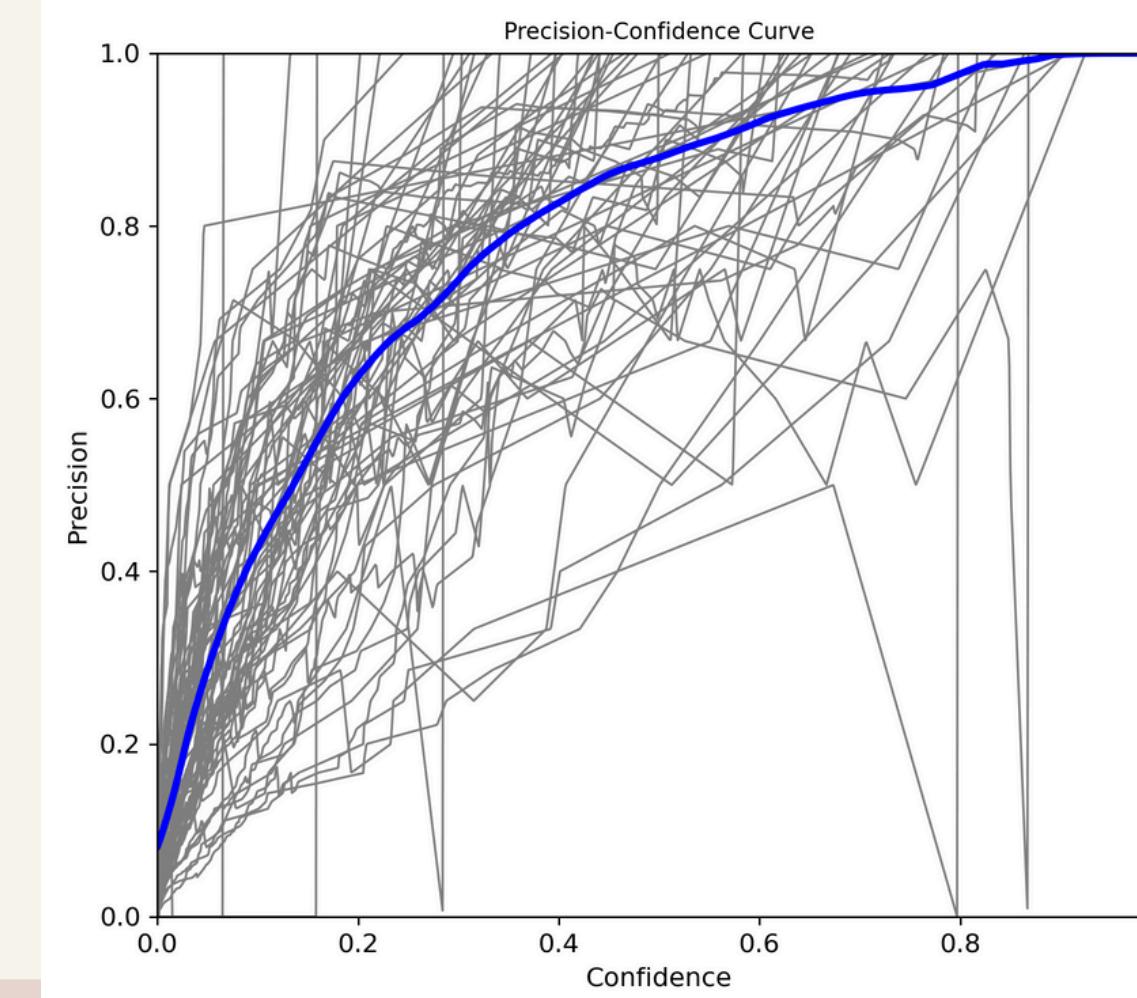
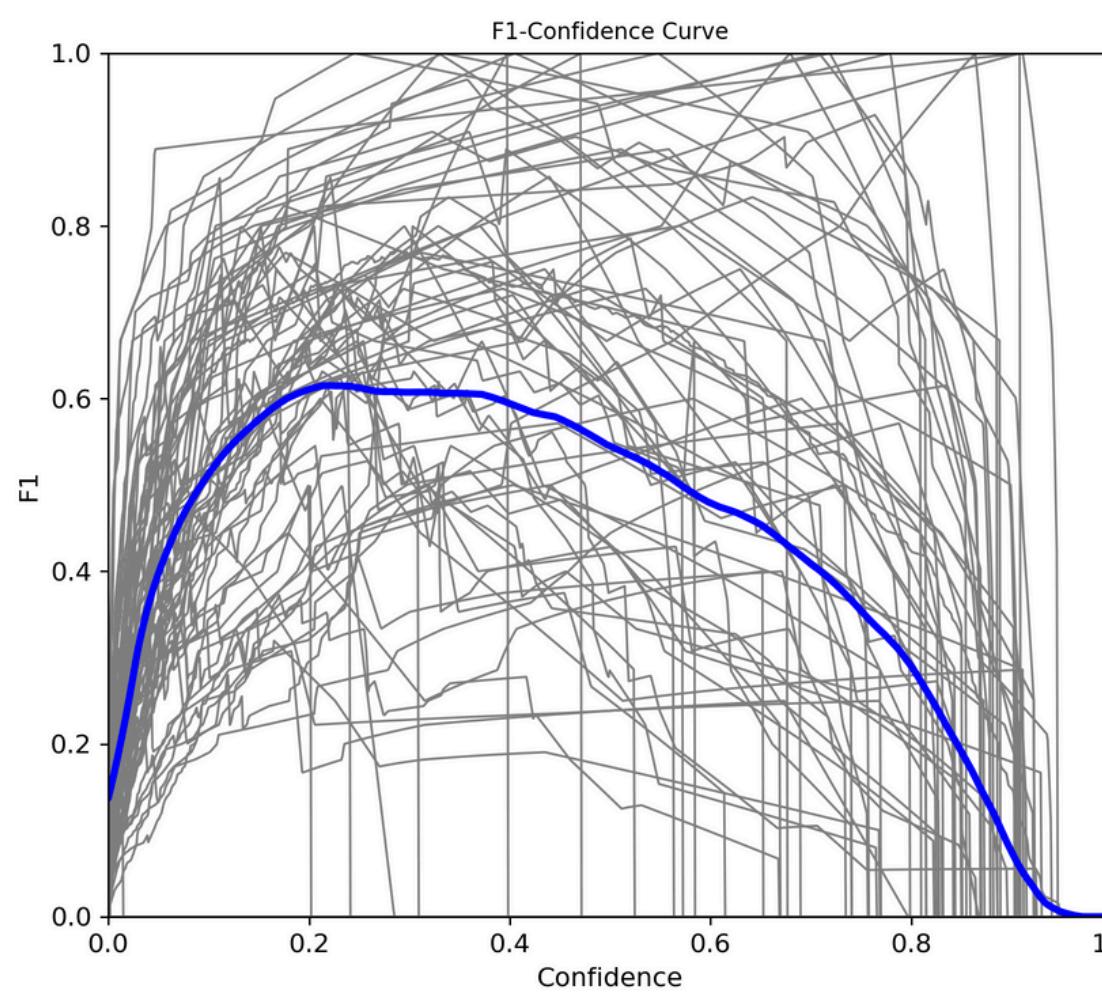
mAP50

0.71

mAP50-95

0.472

Speed: 0.4ms pre-process, 340.0ms inference, 10.9ms NMS/image



MODEL YOLOV5_SAVED_MODEL

Precision

0.669

Recall

0.661

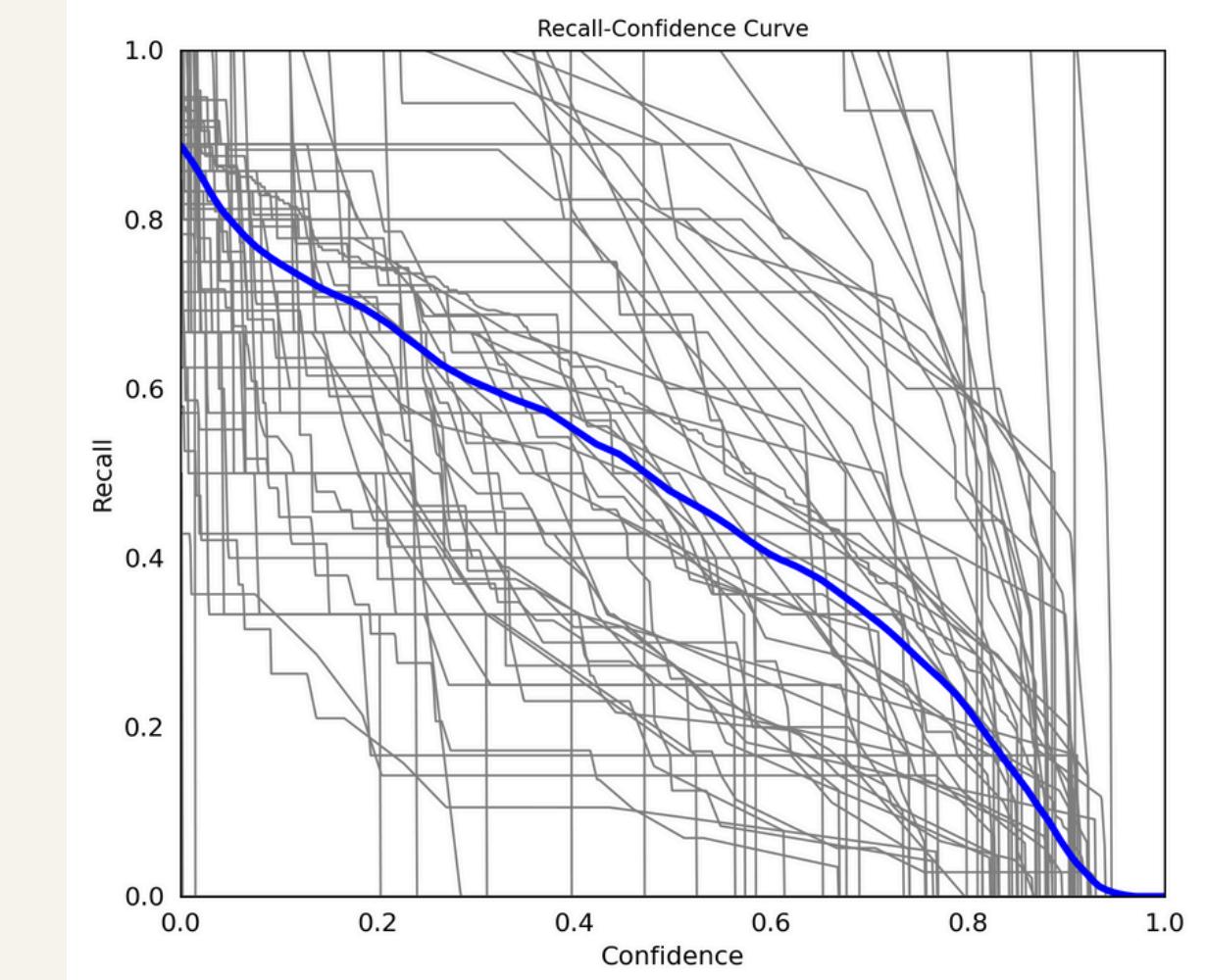
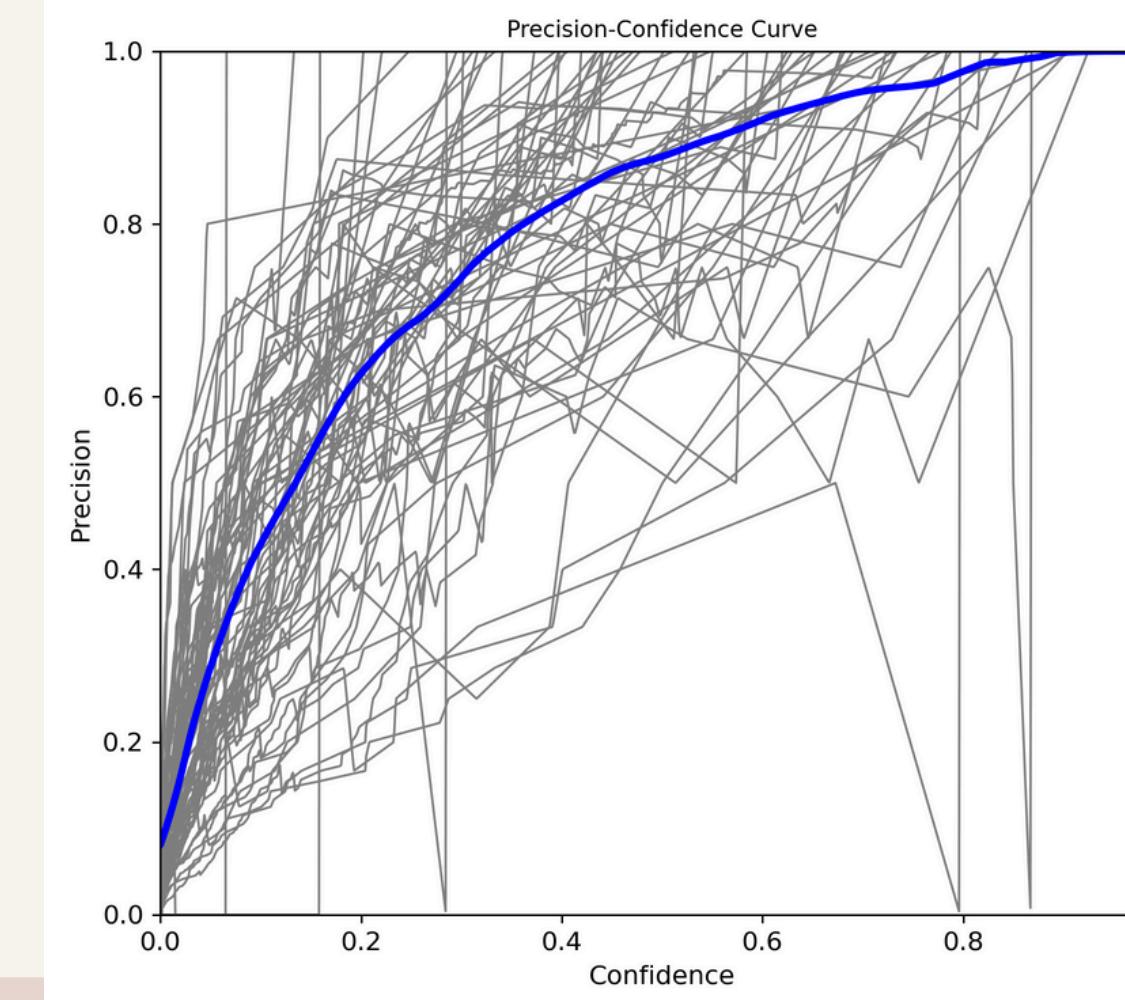
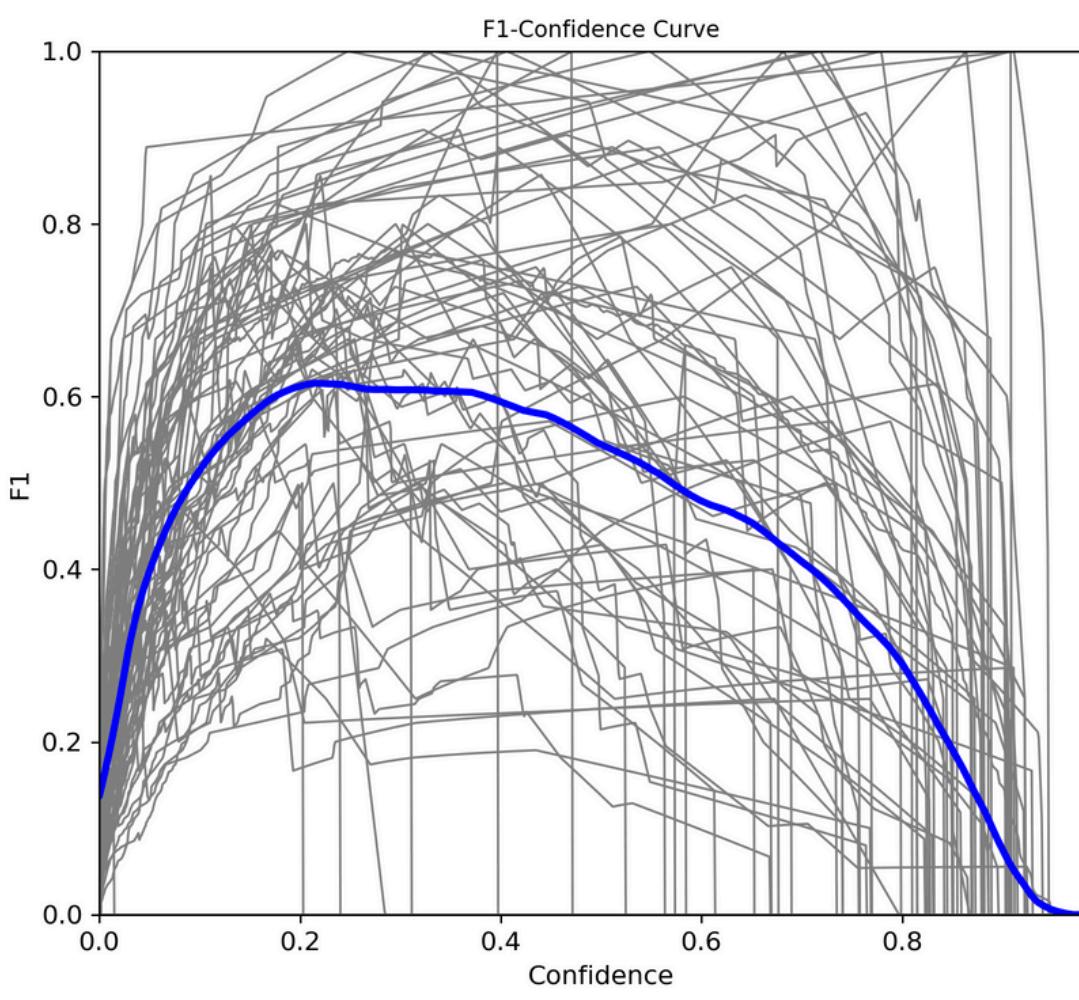
mAP50

0.712

mAP50-95

0.472

Speed: 0.4ms pre-process, 24.5ms inference, 5.6ms NMS/image



MODEL YOLOV5S_OPENVINO_MODEL

Precision

0.669

Recall

0.661

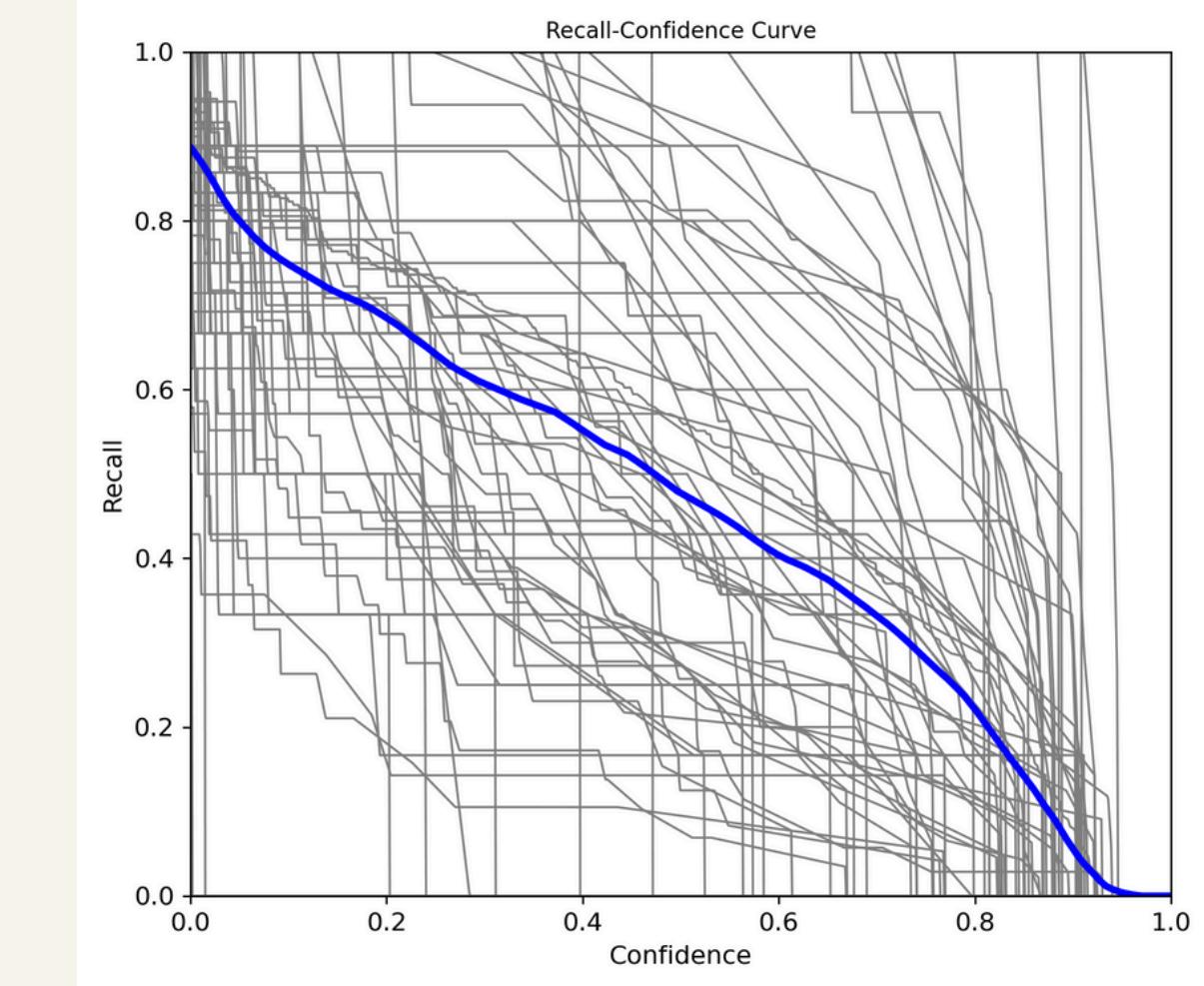
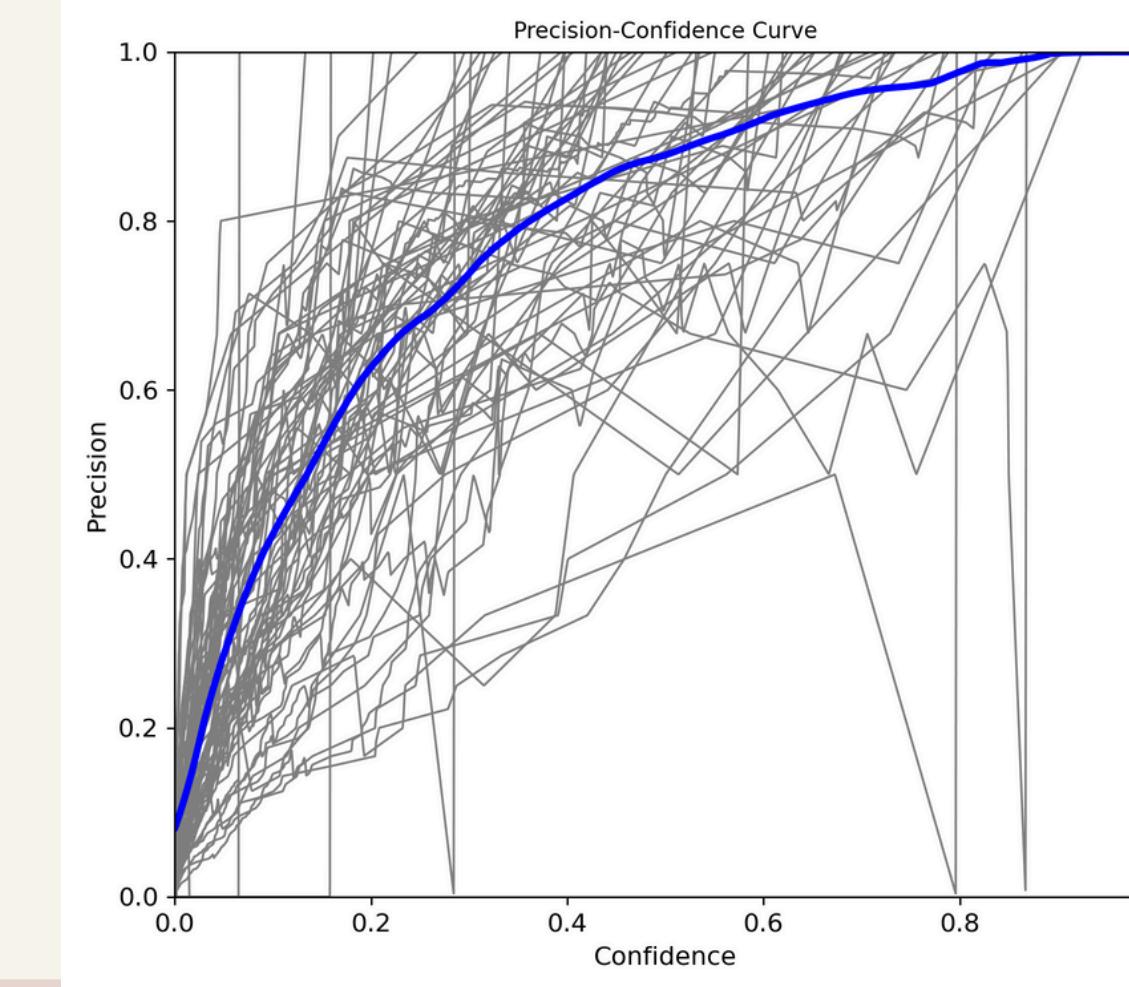
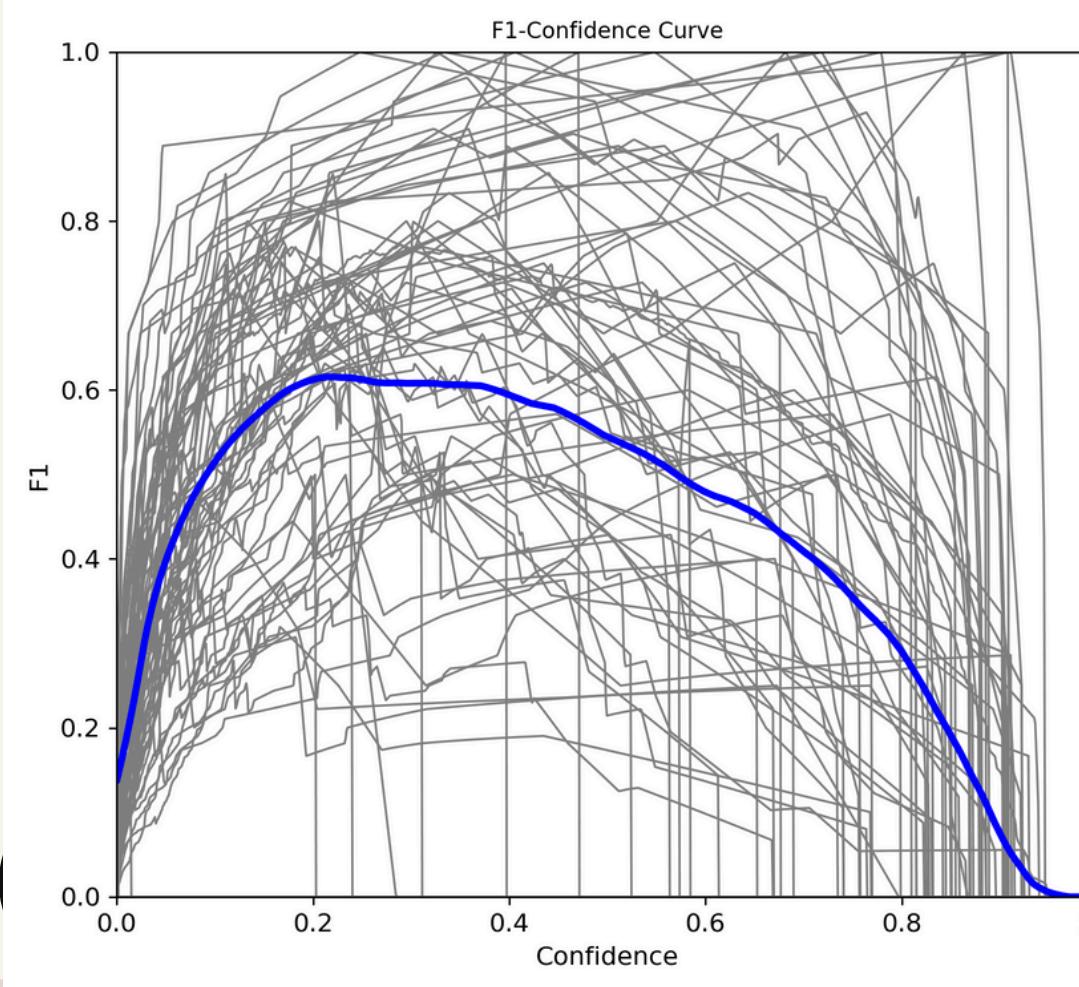
mAP50

0.712

mAP50-95

0.472

Speed: 0.6ms pre-process, 282.3ms inference, 8.0ms NMS/image



SUMMARY

Model	Precision	Recall	mAP50	mAP50-95	Speech
yolov5s.pt (14Mb)	0.671	0.519	0.566	0.371	4.1ms
yolov5s.onnx(28Mb)	0.669	0.661	0.712	0.472	293.0ms
yolov5s.torchscript(28Mb)	0.669	0.661	0.712	0.472	11.7ms
yolov5s-fp16.tflite(14Mb)	0.675	0.657	0.71	0.472	340.0ms
yolov5s_saved_model(28Mb)	0.669	0.661	0.712	0.472	24.5ms
yolov5s_openvino_model(28Mb)	0.669	0.661	0.712	0.472	282.3ms

MODEL PRE-TRAINED

Precision

0.523

Recall

0.335

mAP50

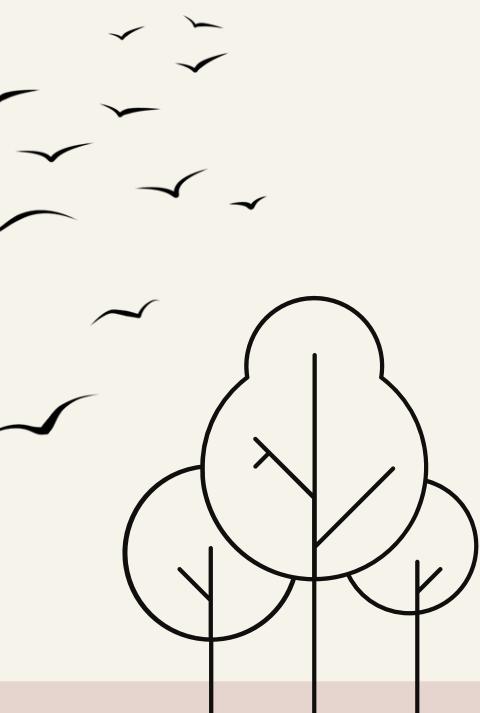
0.34

mAP50-95

0.196

Speed: 0.1ms pre-process, 4.1ms inference, 2.0ms NMS/image

100 epochs completed in 0.195 hours.



MODEL TRAINING FROM SCRATCH

Precision

0.523

Recall

0.335

mAP50

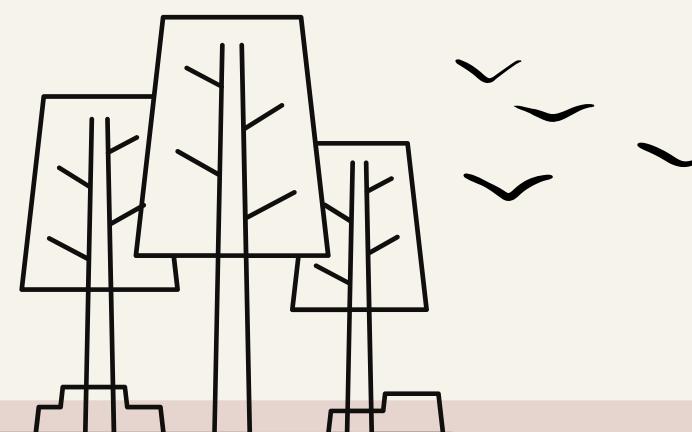
0.34

mAP50-95

0.196

Speed: 0.2ms pre-process, 4.3ms inference, 1.9ms NMS/image

100 epochs completed in 0.196 hours.



Larana University | 2024

THANK YOU

Presented By : Adeline Palmerston