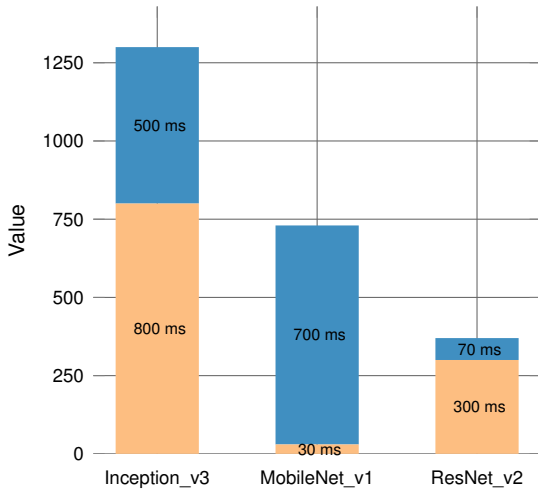
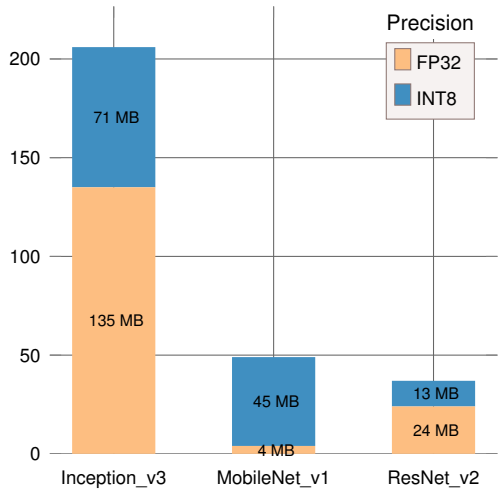


Impact of Quantization on Inference Time and Model Size

Inference_Time



Model_Size



Precision

FP32

INT8

Model