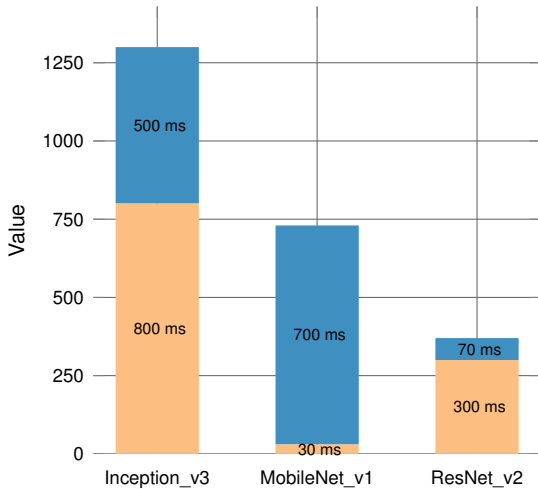
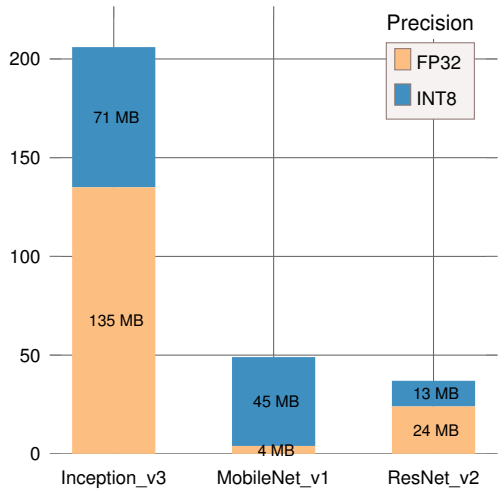


# Impact of Quantization on Inference Time and Model Size

## Inference\_Time



## Model\_Size



Precision

FP32

INT8

Model