resnet50 Inference Latency vs Batch Size