

Runtime of model inference in hot environment
NVIDIA A100, n=100

Benchmark

resnet50

ERT-SQuaD

