Inference throughput by ONNX Runtime Execution Provider

Median seconds per volume

Execution Provider	Value
CPU	5.30
CUDA	0.80
DirectML	1.20