

Inference Latency vs. Network Bandwidth (ViT-large)

