

Impact of batch size on inference speedups

