

Convolution-part for batch size 2, D=768, 18 layers, L=2¹⁶