Transformer encoder max input gradients