Encoder dropout Word Embedding Scaling Positional Encoding 0.1 $X \longrightarrow [x_1, x_2, ..., x_d_model] \longrightarrow$ X / sqrt(d_model) **Encoder Layers** output

N = 4