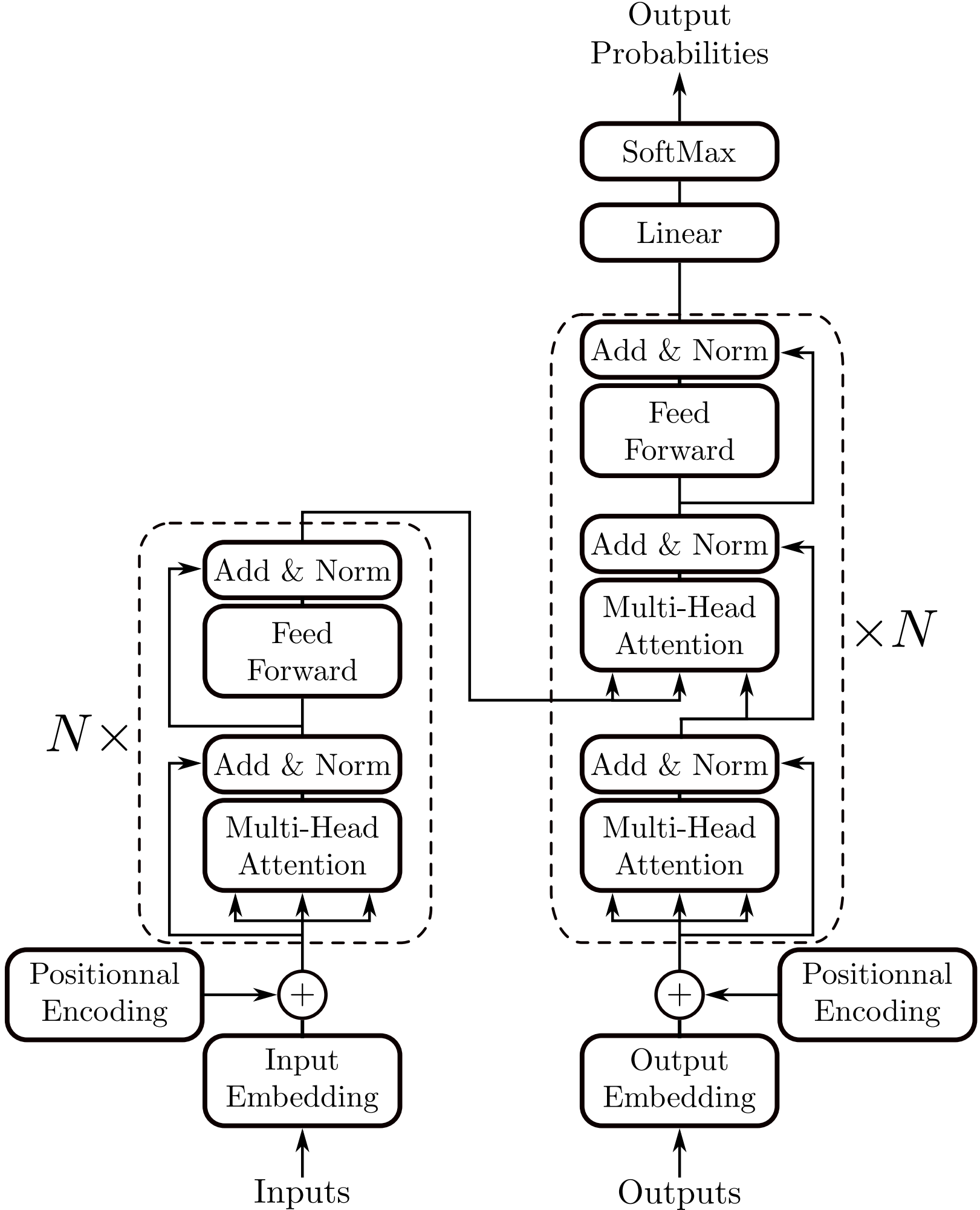
Output
Probabilities
SoftMax
Linear
Add & Norm
Feed
Forward
Add & Norm
Multi-Head
Attention
Add & Norm
Multi-Head
Attention
$\times N$
Add & Norm
Feed
Forward
Add & Norm
Multi-Head
Attention
$N\times$
Positionnal
Encoding
Positionnal
Encoding
Input
Embedding
Output
Embedding
Inputs
Outputs