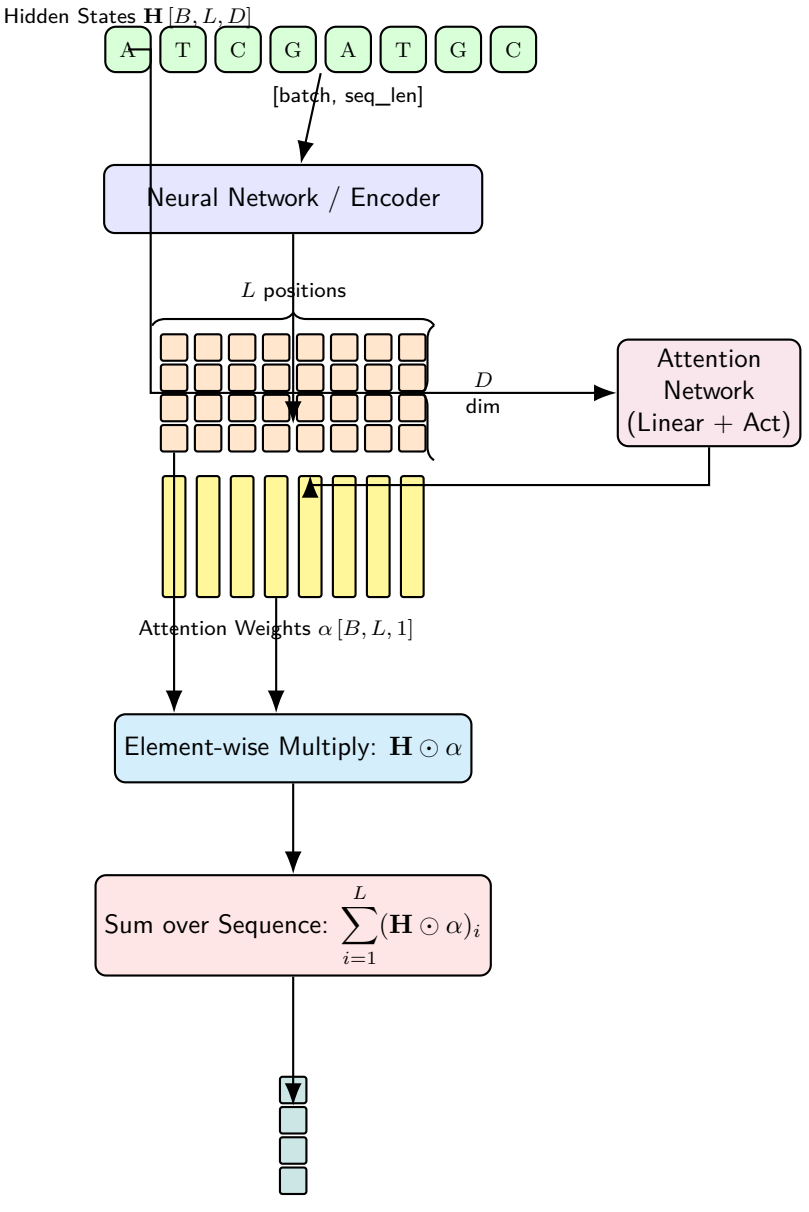
Hidden States $\mathbf{H}\,[B, L, D]$
A
T
C
G
A
T
G
C
[batch, seq_len]
Neural Network / Encoder
L positions
D
dim
Attention
Network
(Linear + Act)
Attention Weights $\alpha\,[B, L, 1]$
Element-wise Multiply: $\mathbf{H} \odot \alpha$
Sum over Sequence: $\sum_{i=1}^{L} (\mathbf{H} \odot \alpha)_i$
Pooled Output $\mathbf{h}_{pool}\,[B, D]$