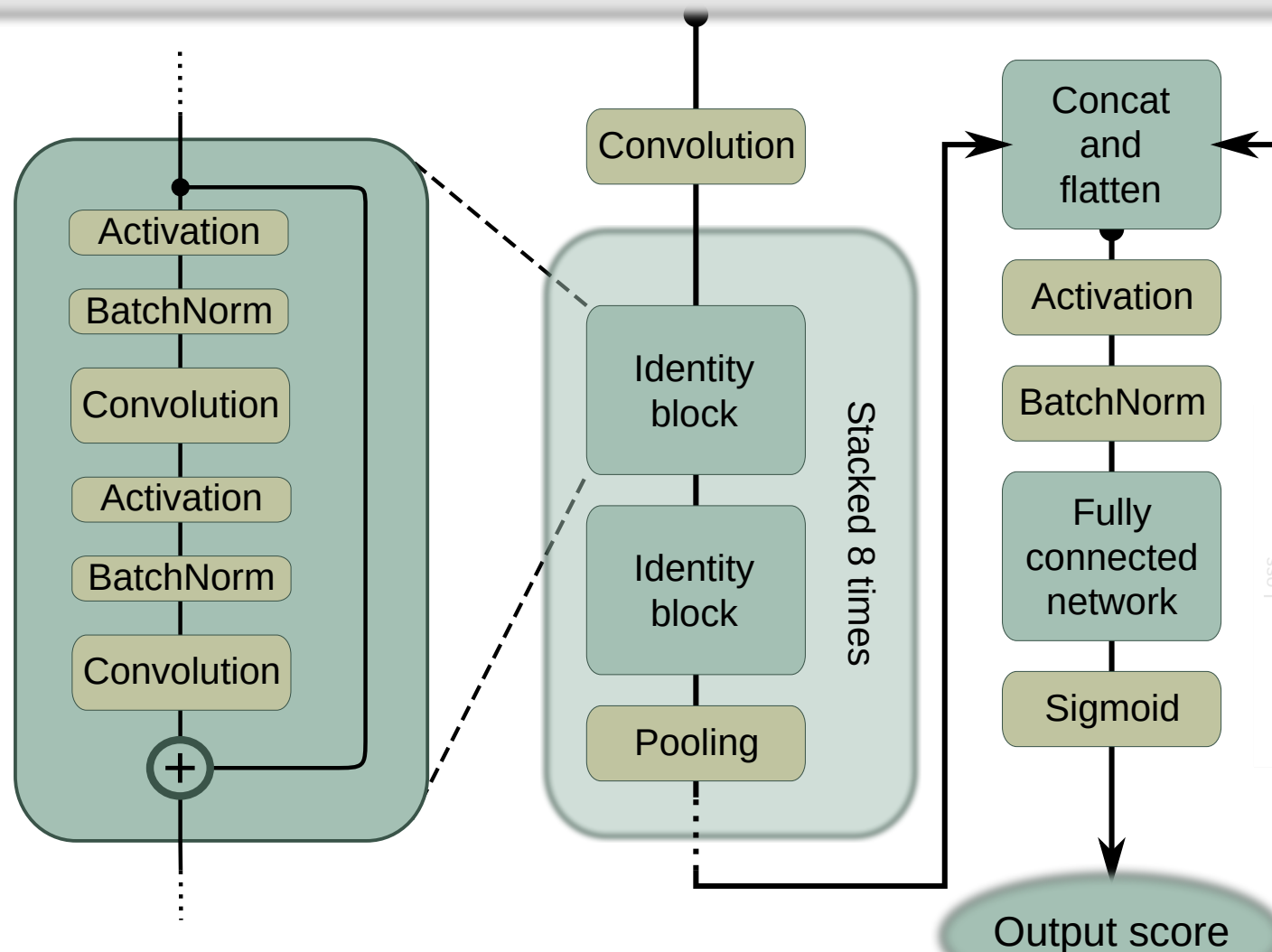


$$dG = \frac{-MFE}{L}$$

Sequence stability



Focal loss instead of cross entropy

$$CE(p_e) = -\log(1 - p_e)$$

$$FL(p_e) = -p_e \log(1 - p_e)$$

