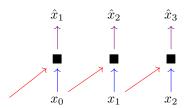
autoregressive network

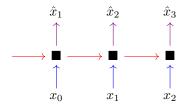
$$x_{0...i} \rightarrow x_{i+1}$$



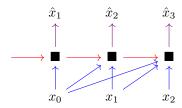
 ${\rm convolution}$



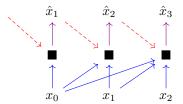
recurrent



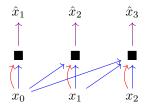
recurrent attention



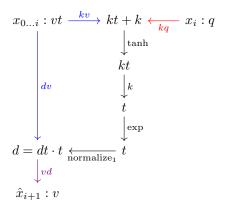
recurrent attention redirected



${\it transformer}$



additive attention



multiplicative attention

$$x_{0...i} : vt \xrightarrow{kv} kt \cdot {}^{T}k \xleftarrow{kq} x_{i} : q$$

$$\downarrow^{dv} \qquad \qquad \downarrow^{\exp}$$

$$d = dt \cdot t \xleftarrow{\text{normalize}_{1}} t$$

$$\downarrow^{vd}$$

$$\hat{x}_{i+1} : v$$