E9 205 Machine Learning for Signal Processing

Recurrent Networks

05-11-2018

Established 1911

INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

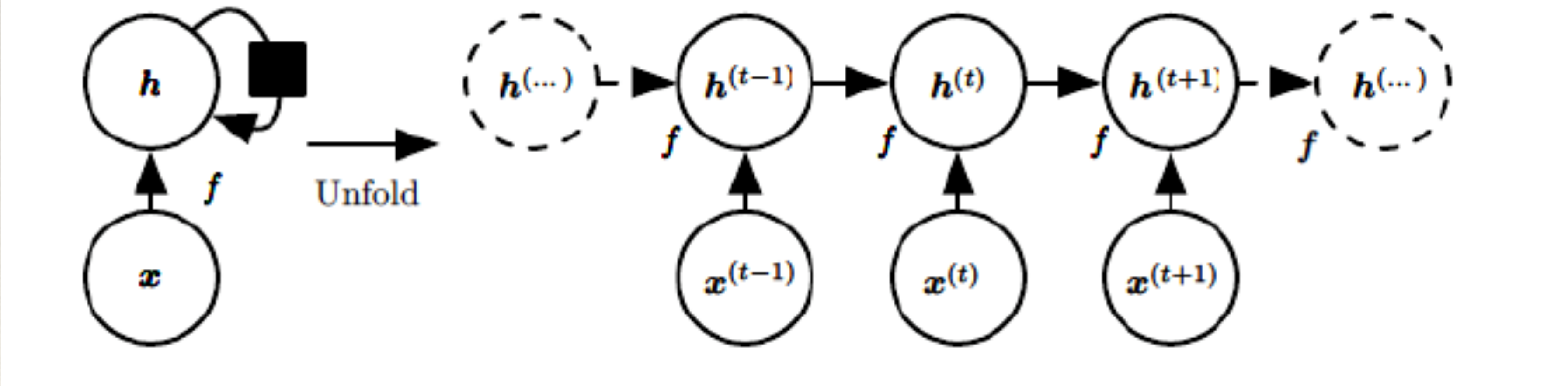
Recurrent Networks

$$s^{(t)} = f(s^{(t-1)}; \theta),$$

$$\begin{aligned} s^{(3)} &= f(s^{(2)}; \theta) \\ &= f(f(s^{(1)}; \theta); \theta) \end{aligned}$$

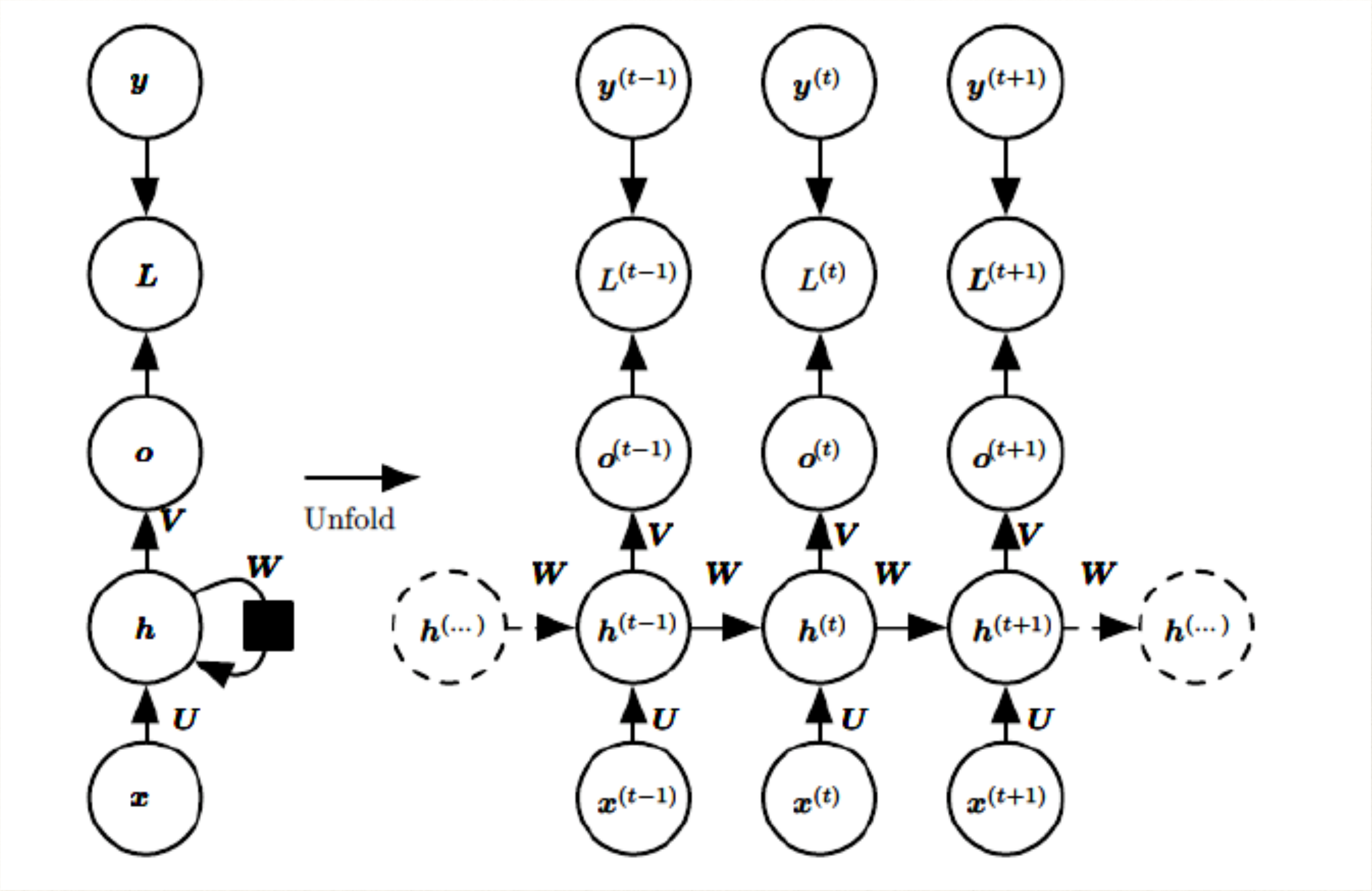
$$s^{(t)} = f(s^{(t-1)}, x^{(t)}; \theta),$$

$$h^{(t)} = f(h^{(t-1)}, x^{(t)}; \theta),$$



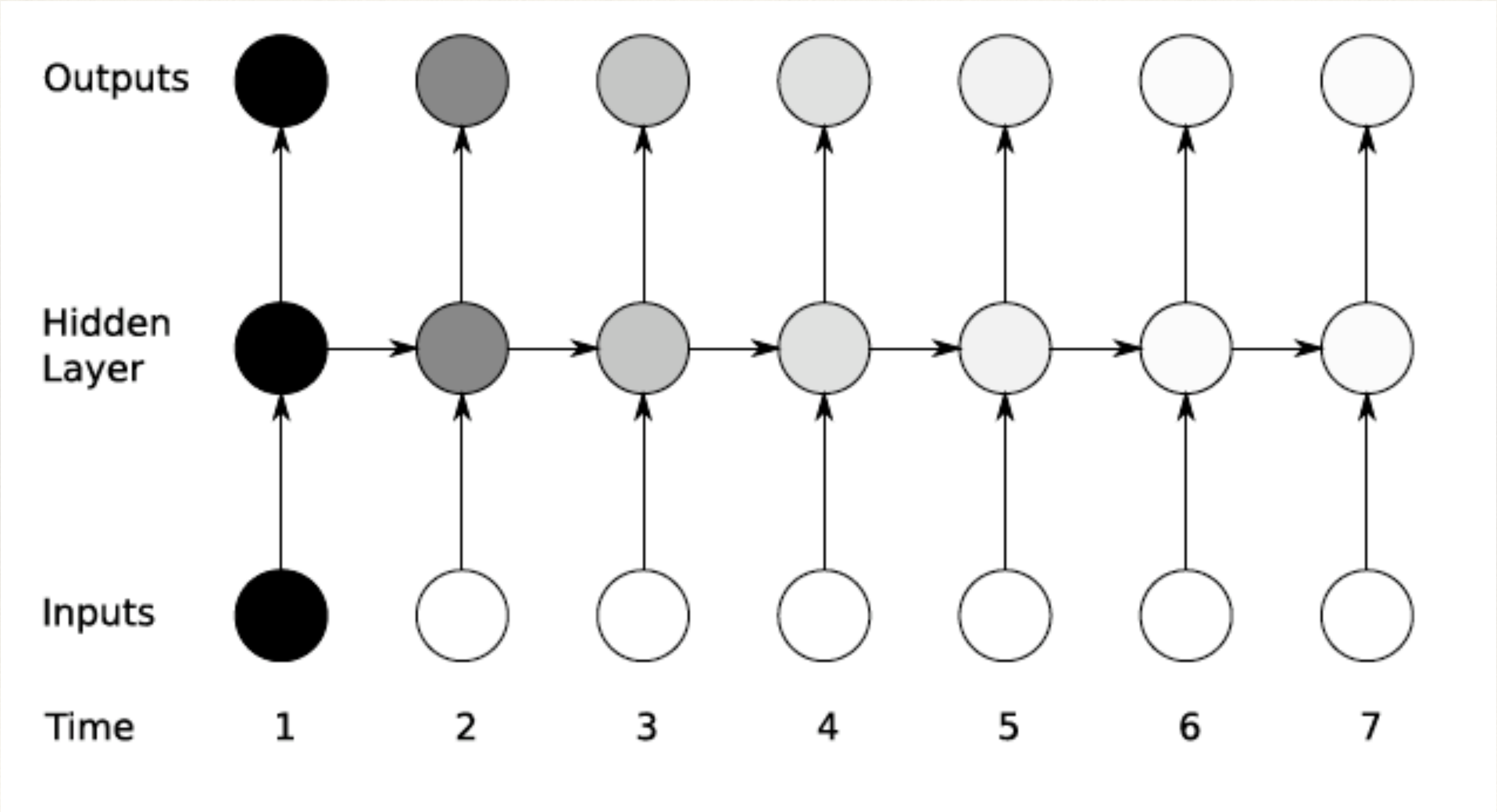
“Deep Learning”, Ian Goodfellow, Yoshua Bengio, Aaron Courville

Recurrent Networks

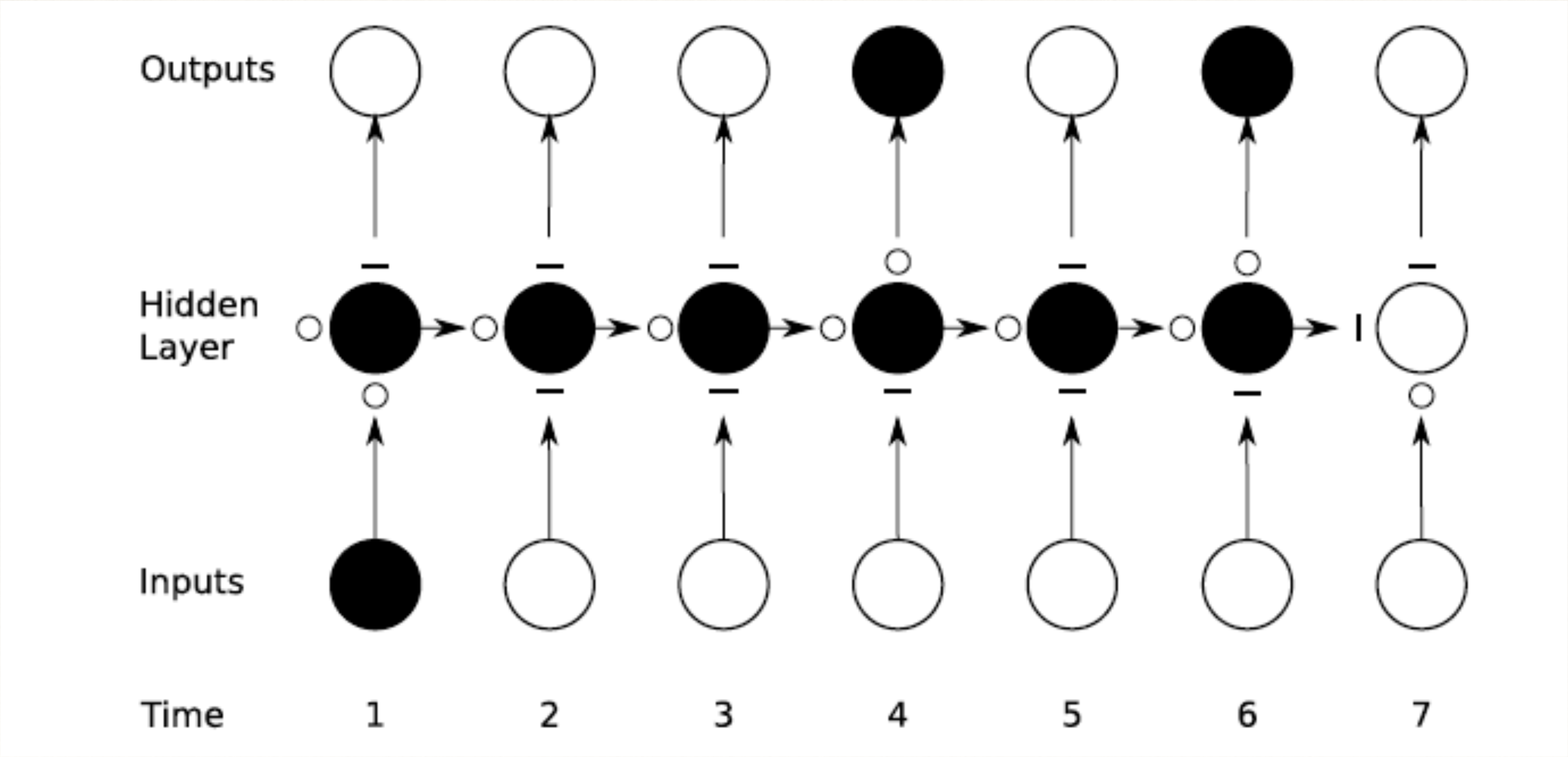


"Deep Learning", Ian Goodfellow, Yoshua Bengio, Aaron Courville

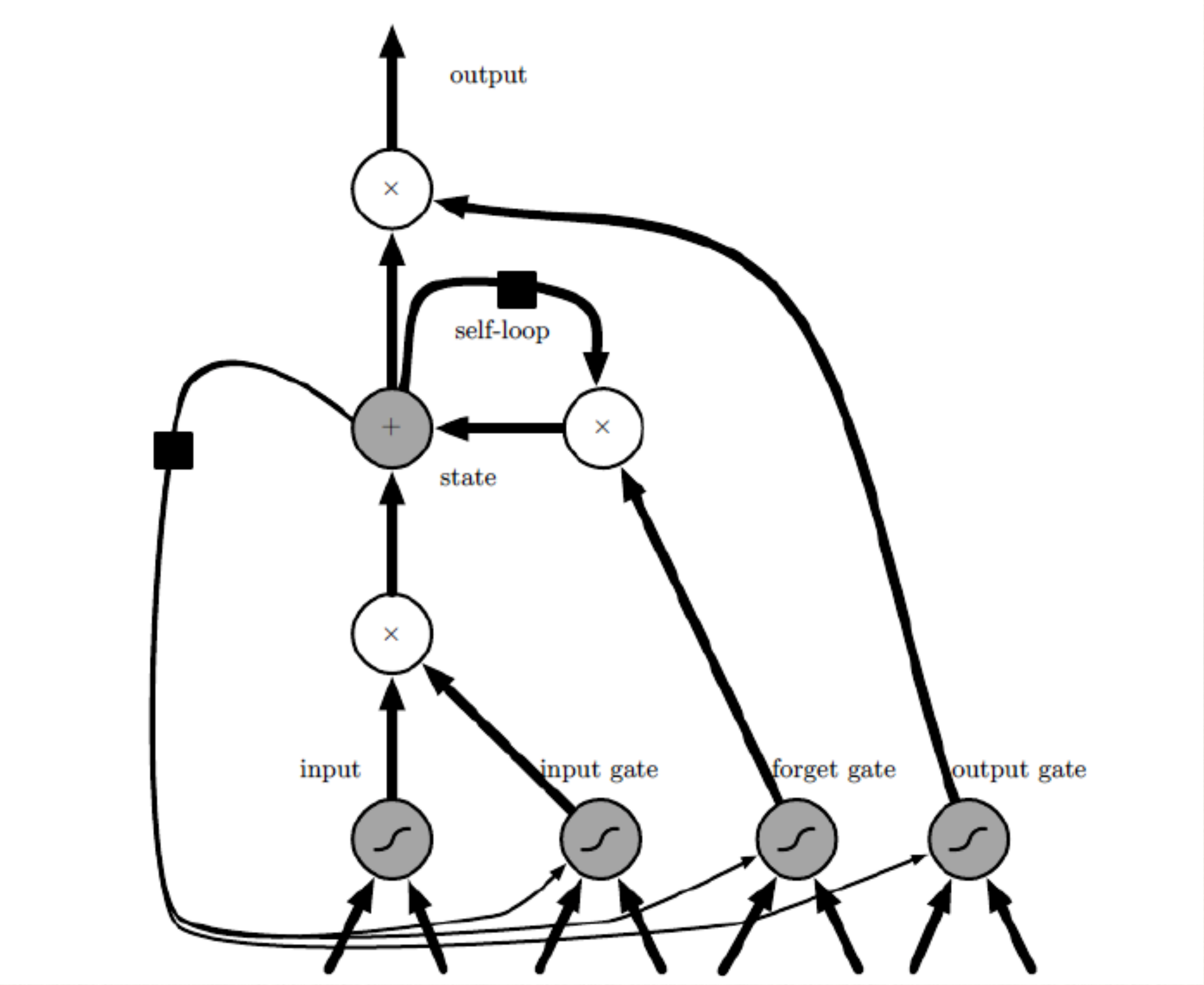
Long-term Dependency Issues



Long-Short Term Memory



Long Short Term Memory Networks



Long Short Term Memory Networks

$$f_i^{(t)} = \sigma\left(b_i^f + \sum_j U_{i,j}^f x_j^{(t)} + \sum_j W_{i,j}^f h_j^{(t-1)}\right)$$

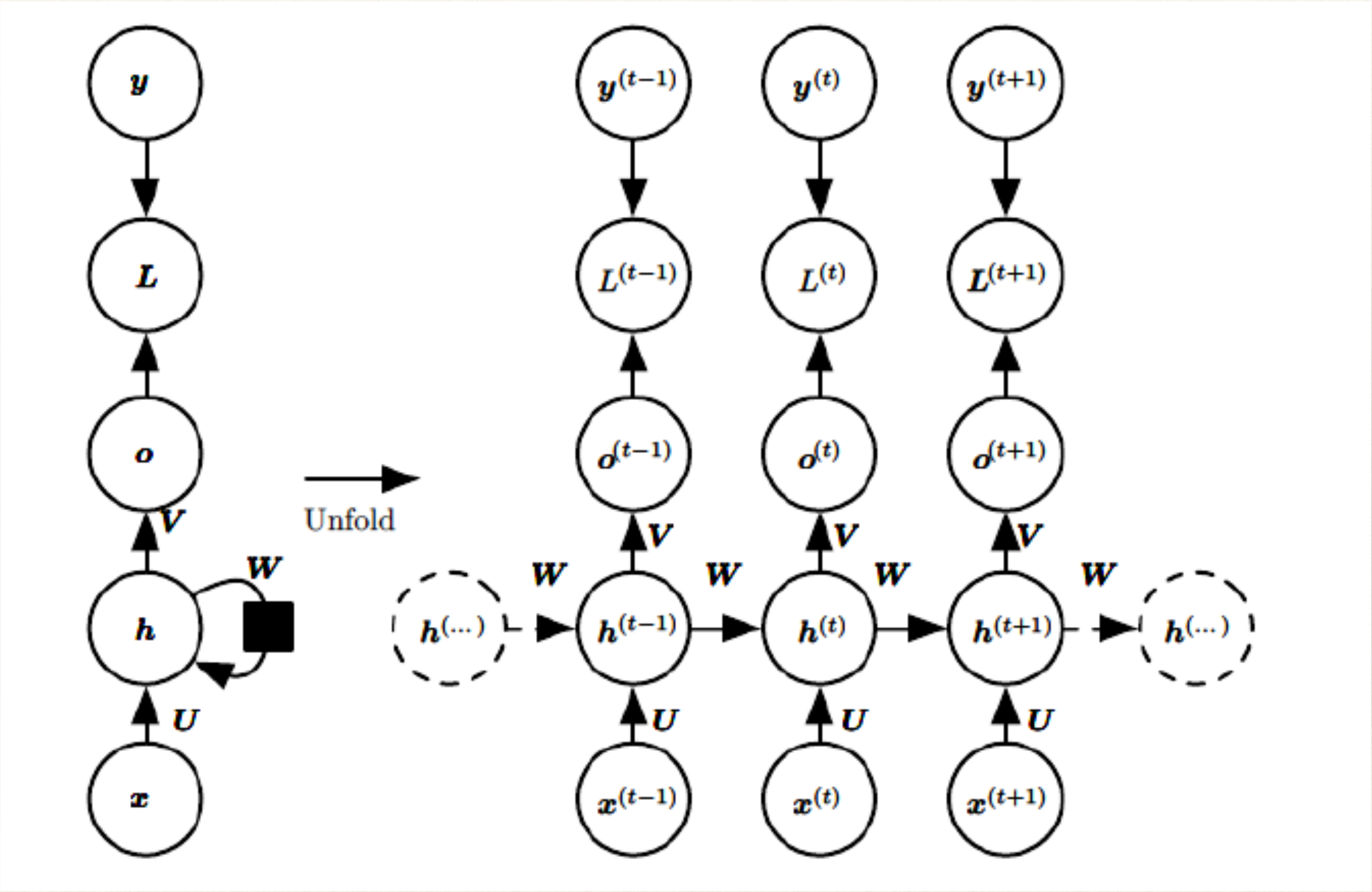
$$s_i^{(t)} = f_i^{(t)} s_i^{(t-1)} + g_i^{(t)} \sigma\left(b_i + \sum_j U_{i,j} x_j^{(t)} + \sum_j W_{i,j} h_j^{(t-1)}\right)$$

$$g_i^{(t)} = \sigma\left(b_i^g + \sum_j U_{i,j}^g x_j^{(t)} + \sum_j W_{i,j}^g h_j^{(t-1)}\right)$$

$$h_i^{(t)} = \tanh\left(s_i^{(t)}\right) q_i^{(t)}$$

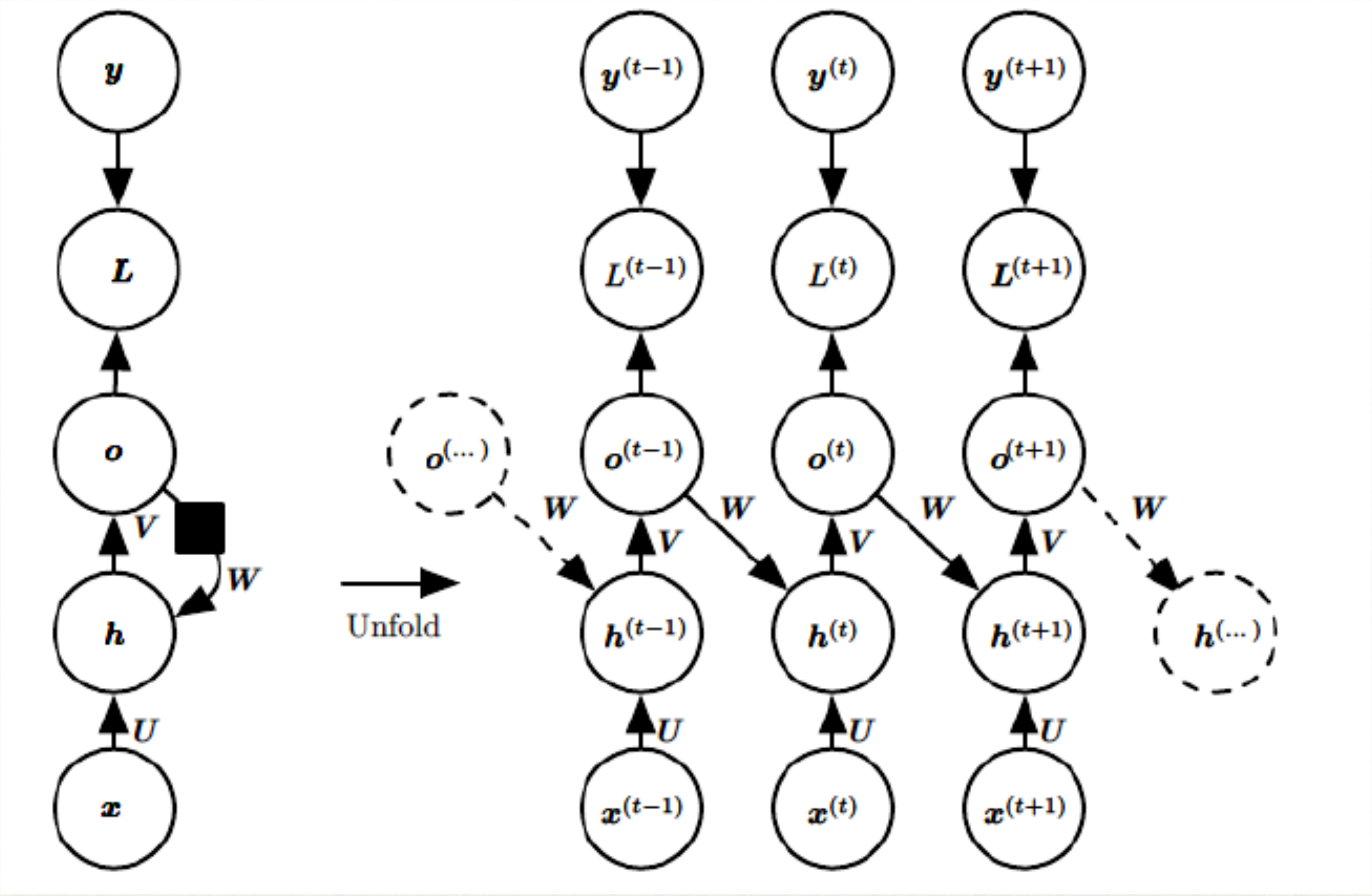
$$q_i^{(t)} = \sigma\left(b_i^o + \sum_j U_{i,j}^o x_j^{(t)} + \sum_j W_{i,j}^o h_j^{(t-1)}\right)$$

Standard Recurrent Networks



"Deep Learning", Ian Goodfellow, Yoshua Bengio, Aaron Courville

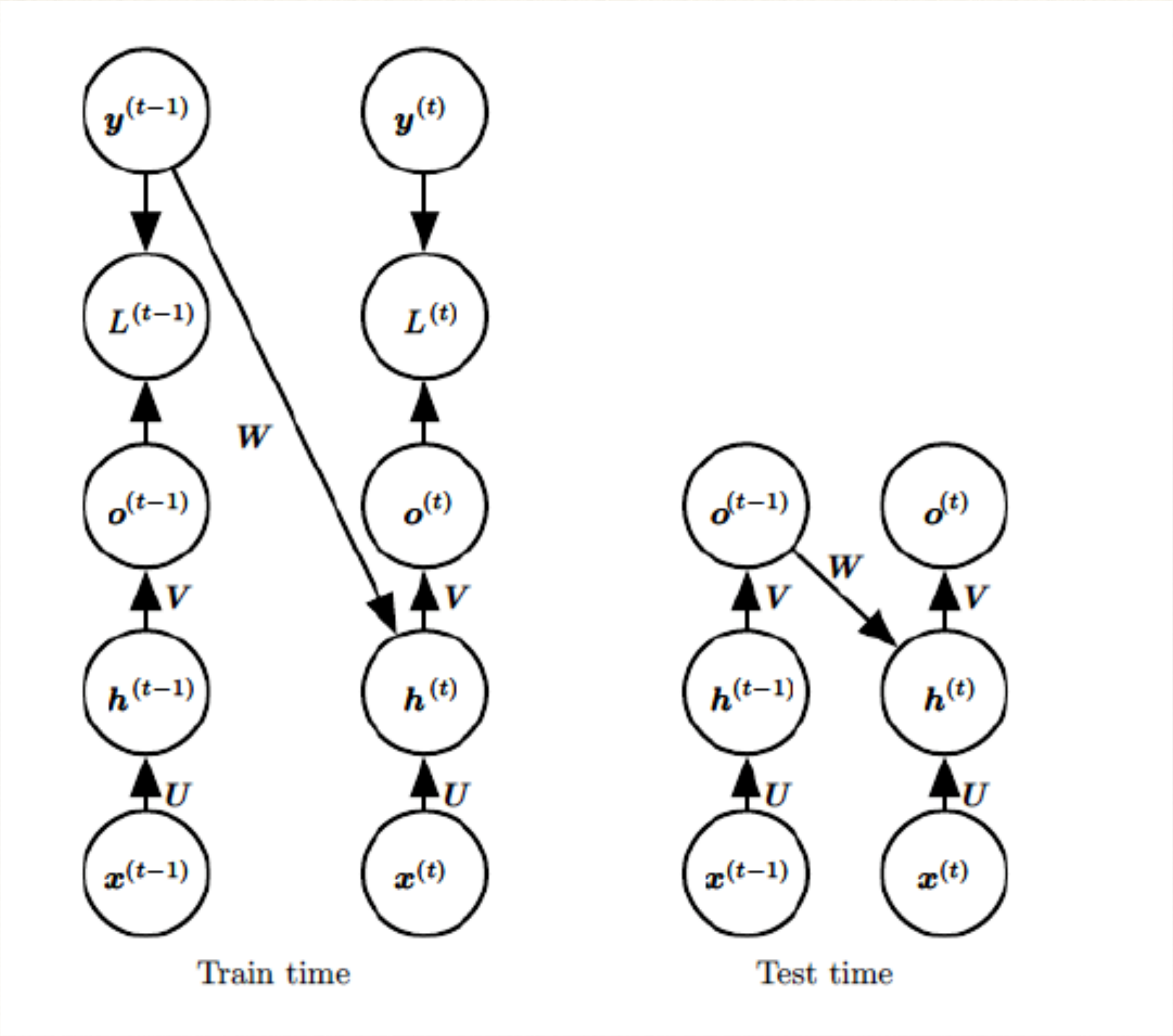
Other Recurrent Networks



Teacher Forcing Networks

"Deep Learning", Ian Goodfellow, Yoshua Bengio, Aaron Courville

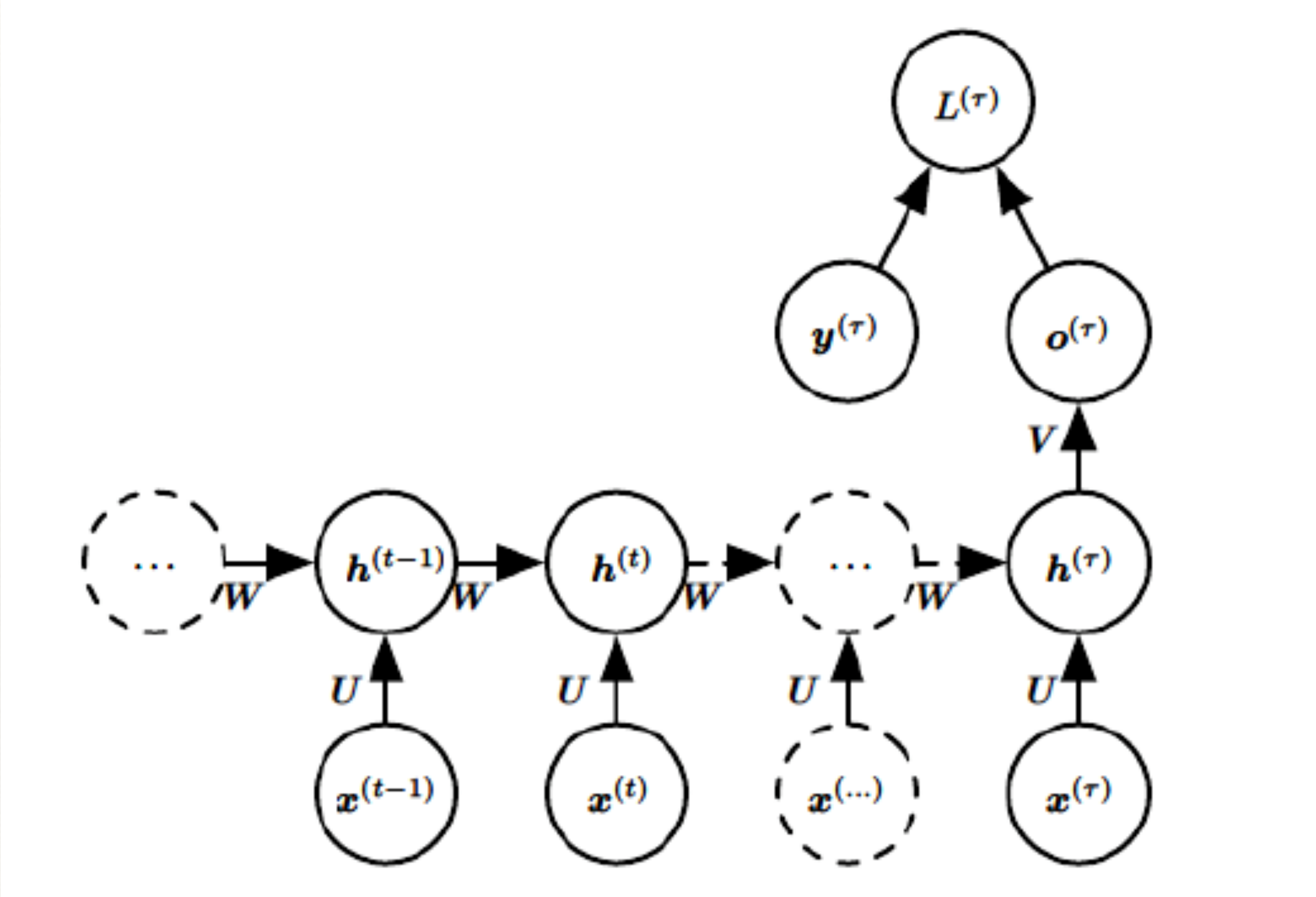
Recurrent Networks



Teacher Forcing Networks

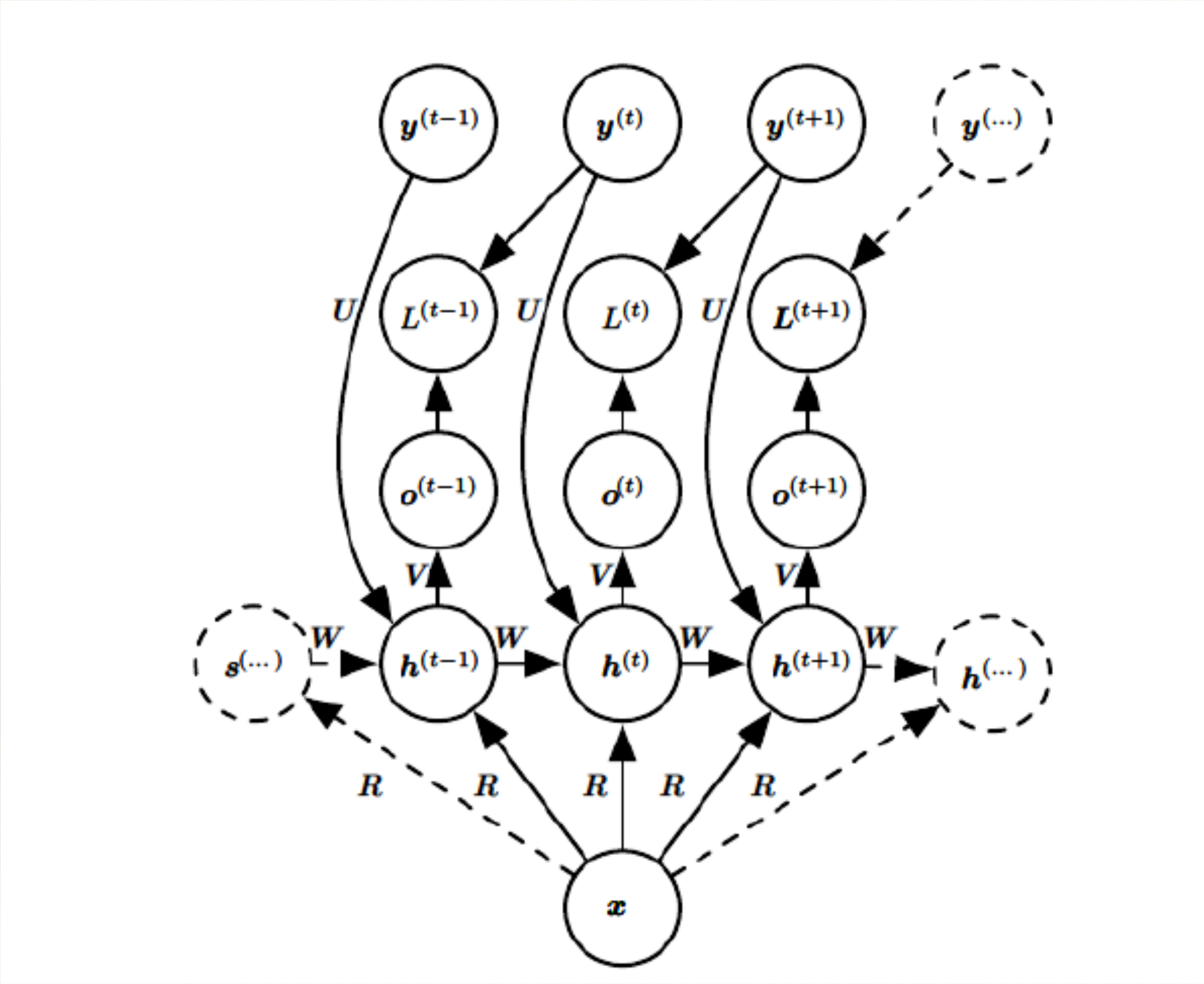
"Deep Learning", Ian Goodfellow, Yoshua Bengio, Aaron Courville

Recurrent Networks



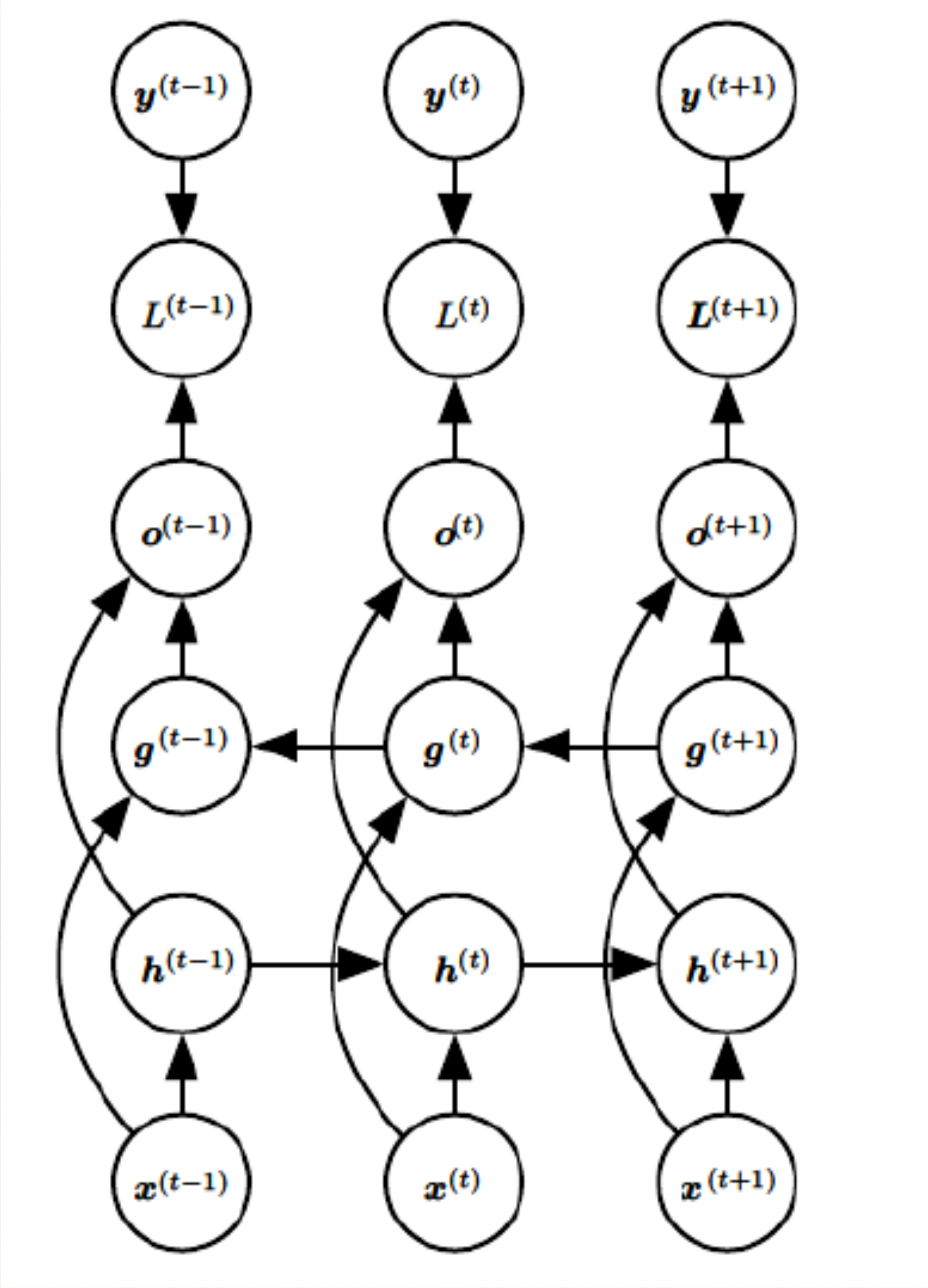
Multiple Input Single Output

Recurrent Networks



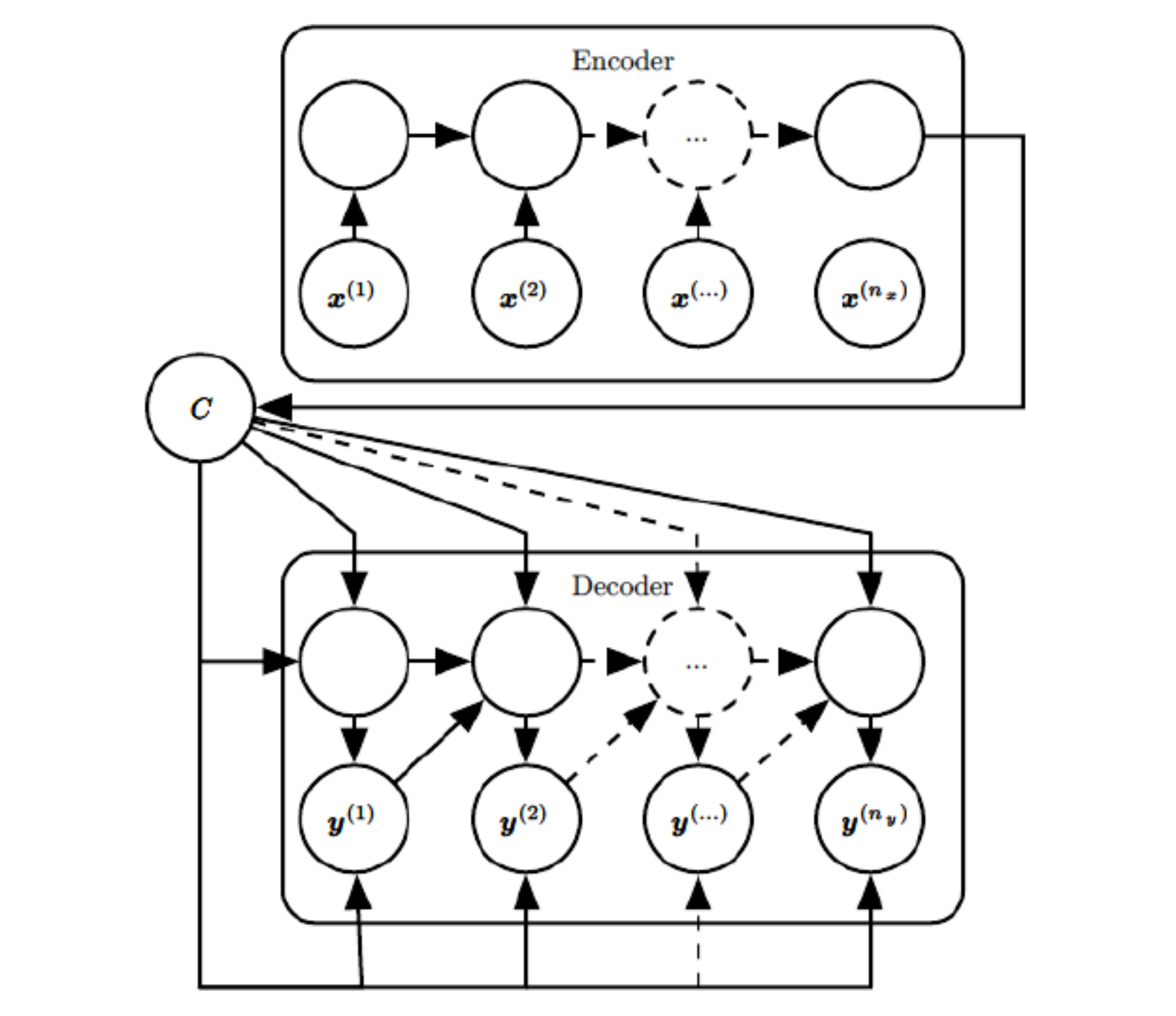
Single Input
Multiple Output

Recurrent Networks



Bi-directional Networks

Recurrent Networks



Sequence to Sequence Mapping Networks