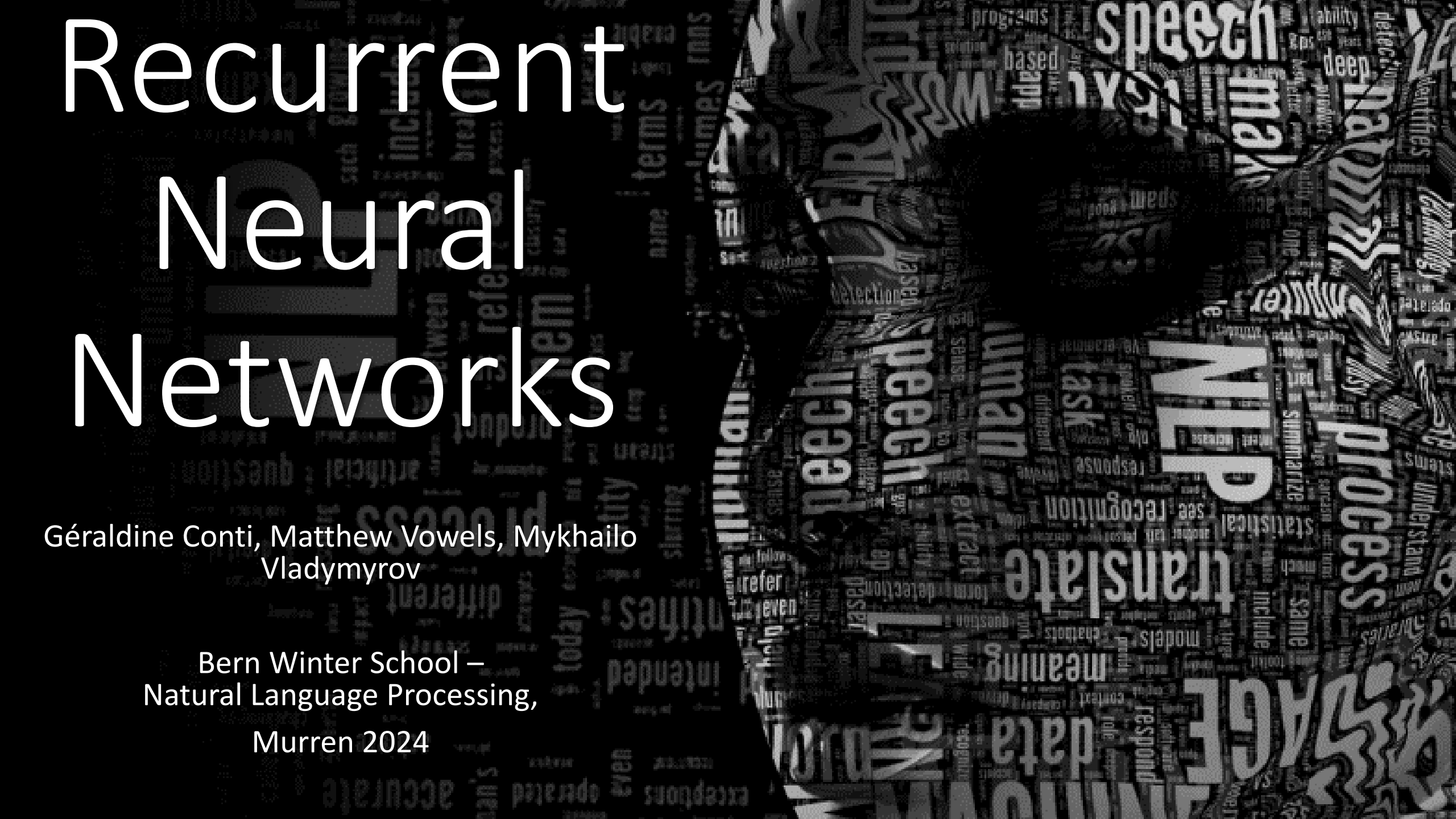
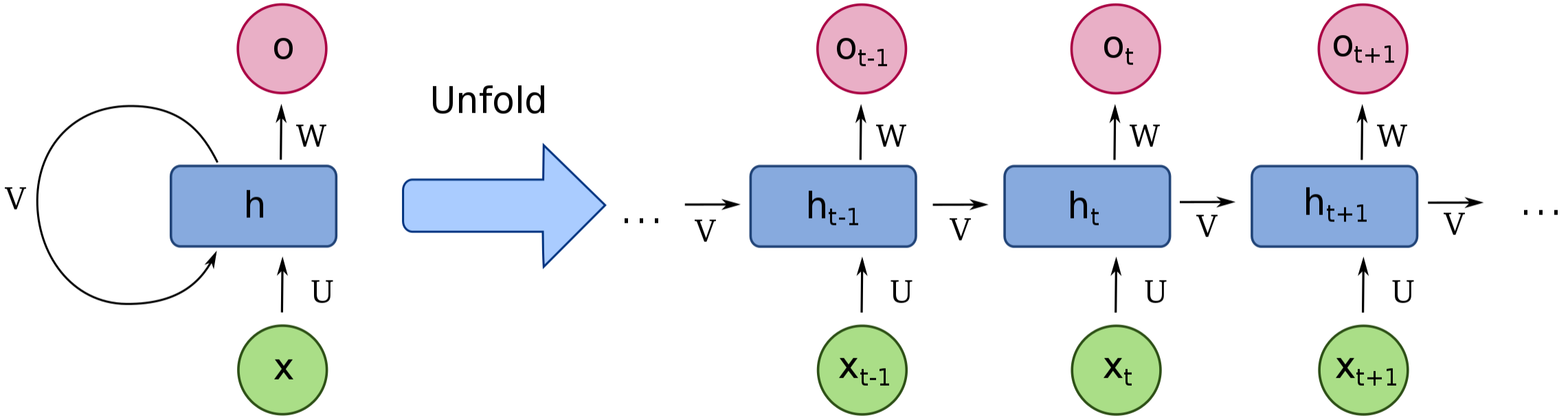
Recurrent Neural Networks



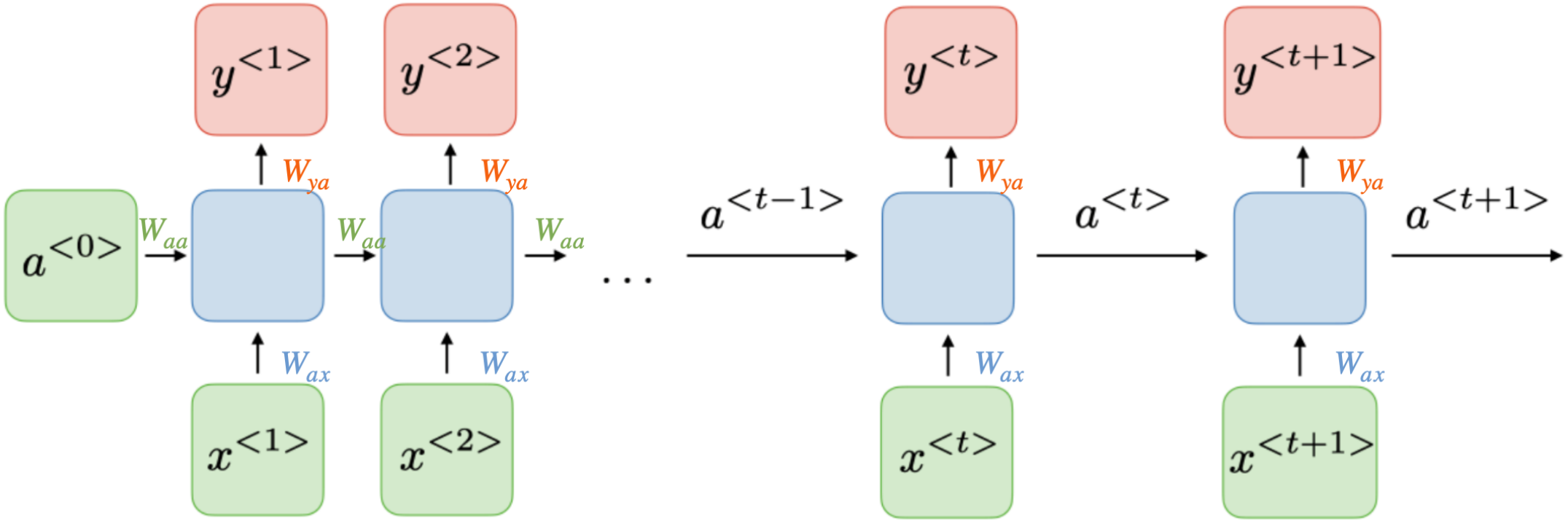
Géraldine Conti, Matthew Vowels, Mykhailo Vladymyrov

Bern Winter School –
Natural Language Processing,
Murren 2024

RNN model



RNN model

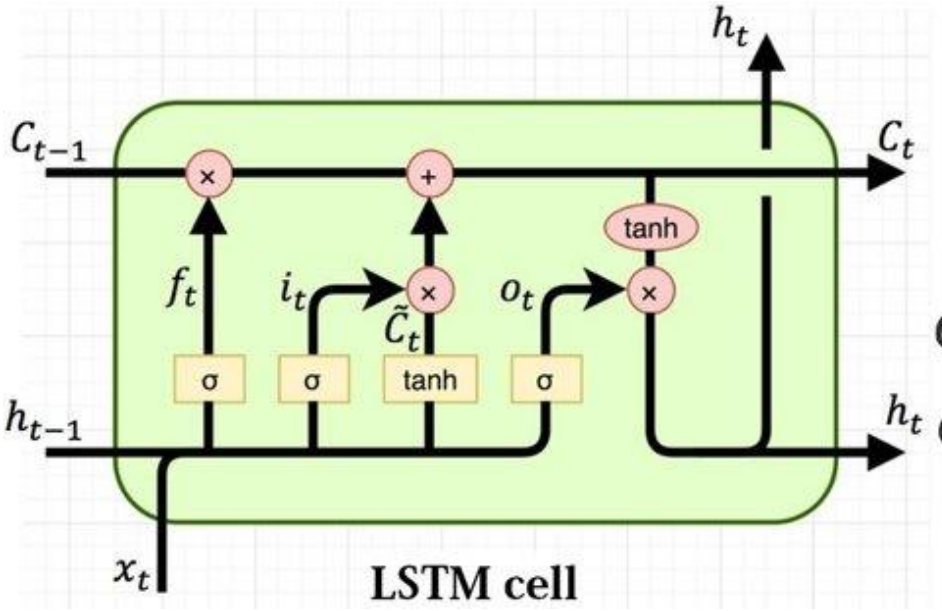


$$a^{<t>} = g_1(W_{aa}a^{<t-1>} + W_{ax}x^{<t>} + b_a)$$

$$y^{<t>} = g_2(W_{ya}a^{<t>} + b_y)$$

W_{ax} ,W_{aa}, W_{ya}, b_a and b_y are weights that are shared temporally and g_1, g_2 activation functions

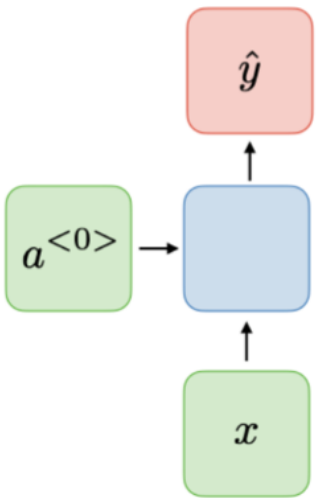
LSTM model



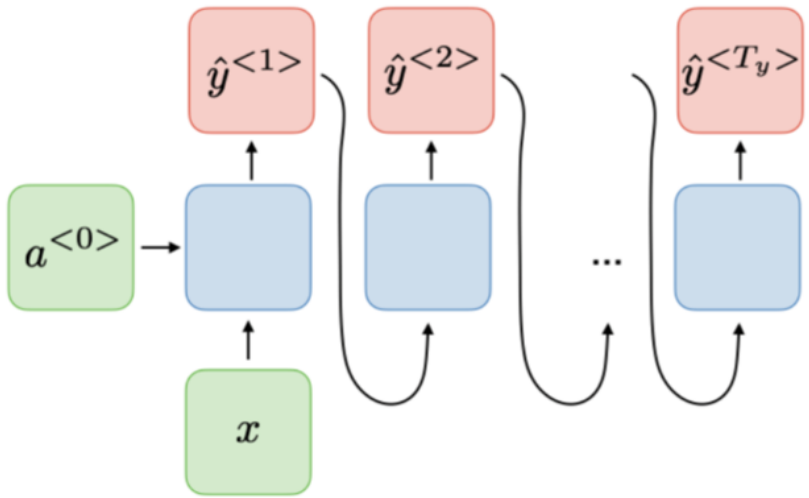
$$i_t = \sigma\left(x_t U^i + h_{t-1} W^i\right)$$
$$f_t = \sigma\left(x_t U^f + h_{t-1} W^f\right)$$
$$o_t = \sigma\left(x_t U^o + h_{t-1} W^o\right)$$
$$\tilde{C}_t = \tanh\left(x_t U^g + h_{t-1} W^g\right)$$
$$C_t = \sigma\left(f_t * C_{t-1} + i_t * \tilde{C}_t\right)$$
$$h_t = \tanh(C_t) * o_t$$

Varsamopoulos, Savvas & Bertels, Koen & Almudever, Carmen. (2018). Designing neural network based decoders for surface codes.

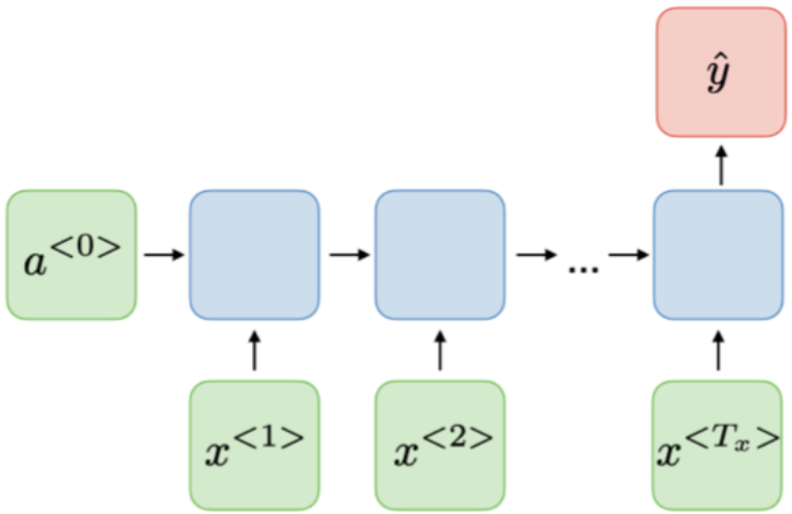
Applications of RNN

Type of RNN	Illustration	Example
One-to-one $T_x = T_y = 1$		Traditional neural network

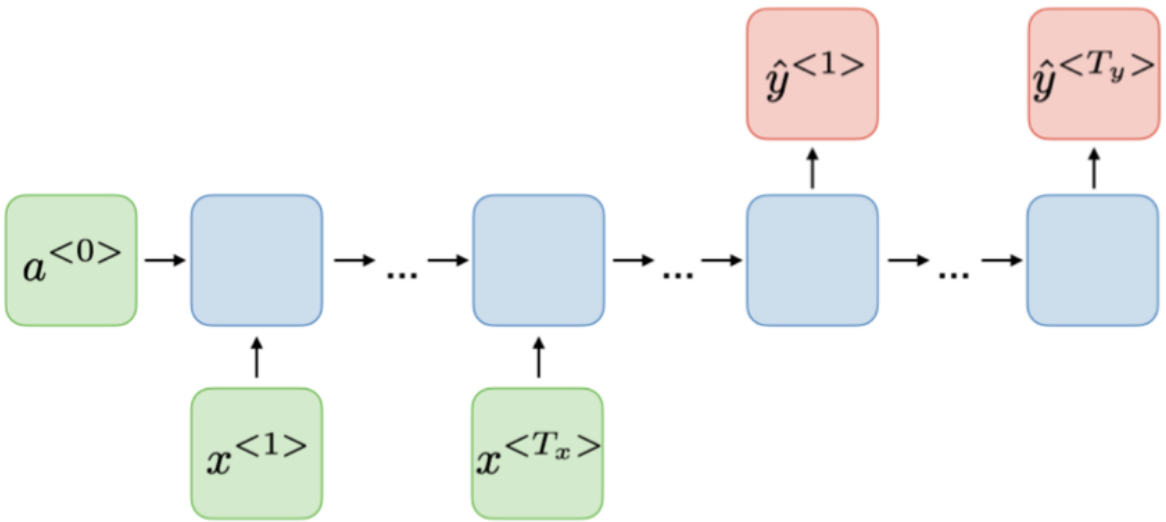
Applications of RNN

Type of RNN	Illustration	Example
One-to-many $T_x = 1, T_y > 1$		Music generation

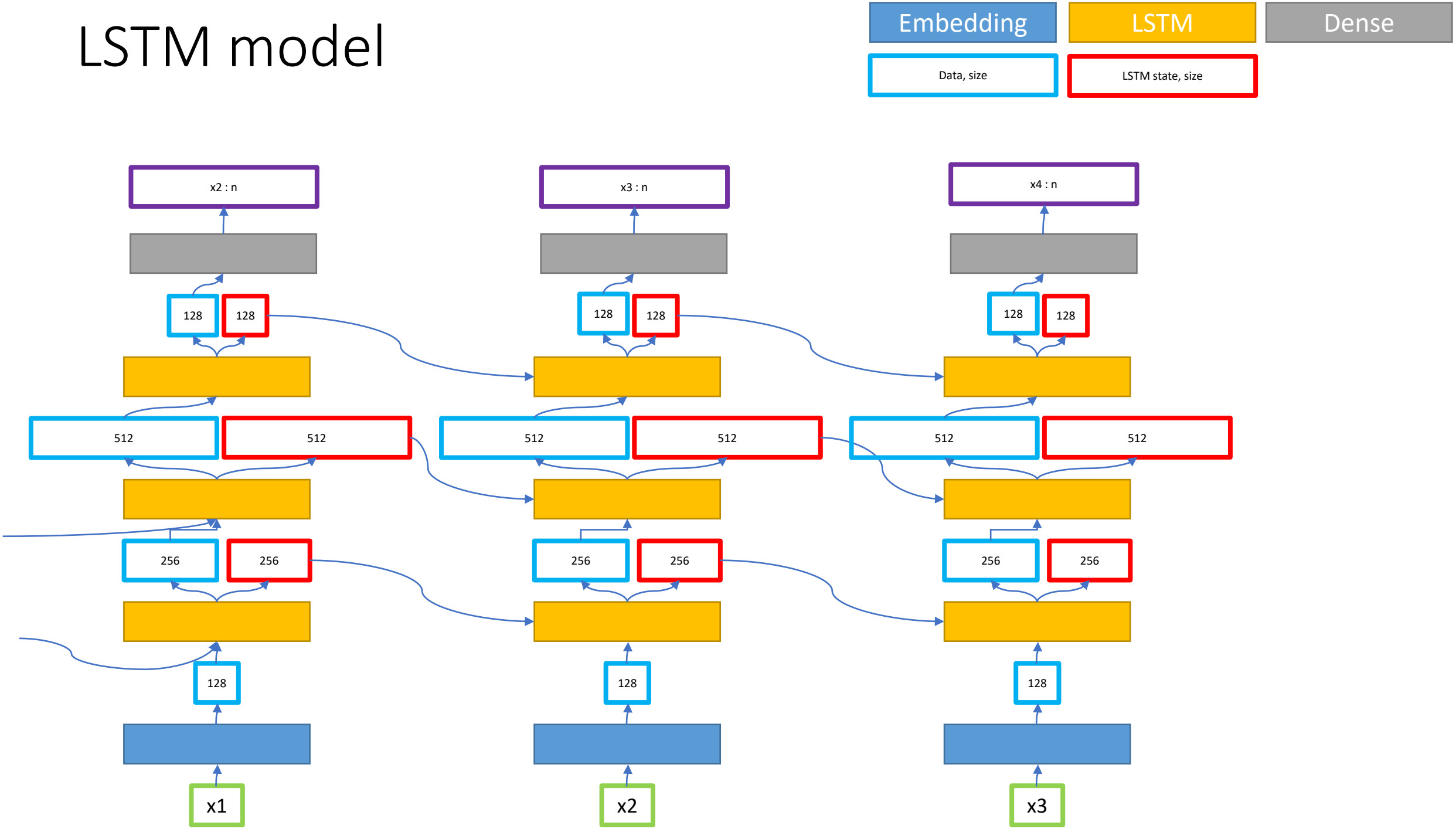
Applications of RNN

Type of RNN	Illustration	Example
Many-to-one $T_x > 1, T_y = 1$		Sentiment classification

Applications of RNN

Type of RNN	Illustration	Example
Many-to-many $T_x \neq T_y$		Machine translation

LSTM model


Embedding
LSTM
Dense
Data, size
LSTM state, size
x2 : n
x3 : n
x4 : n
128
128
512
512
256
256
x1
x2
x3

Key elements of neural networks

- Fully connected (dense) layer
- Drop out
- Normalization
- Residual connection
- Attention mechanism