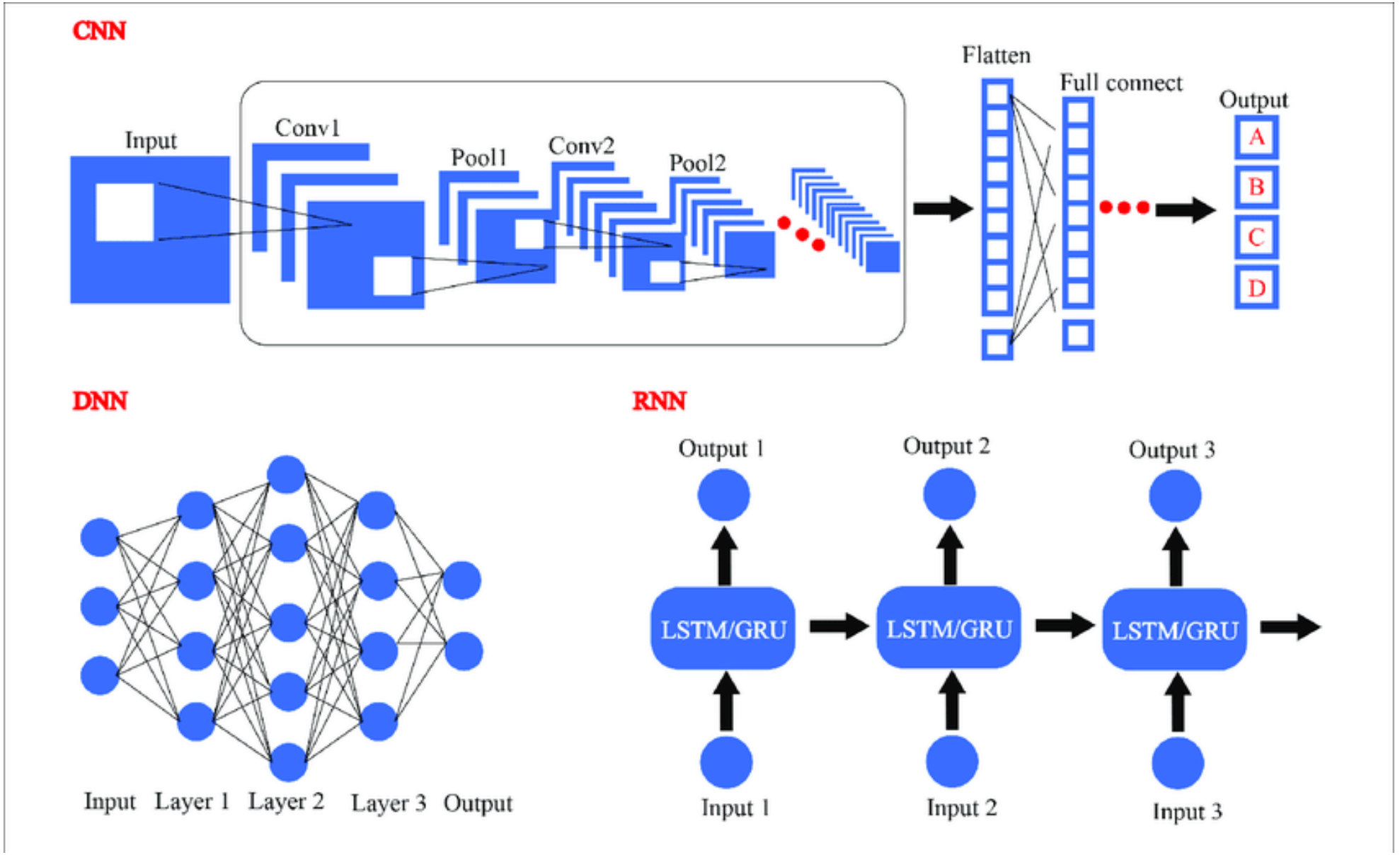
神經與行為模型建構
(Neural & Behavioral Modeling)

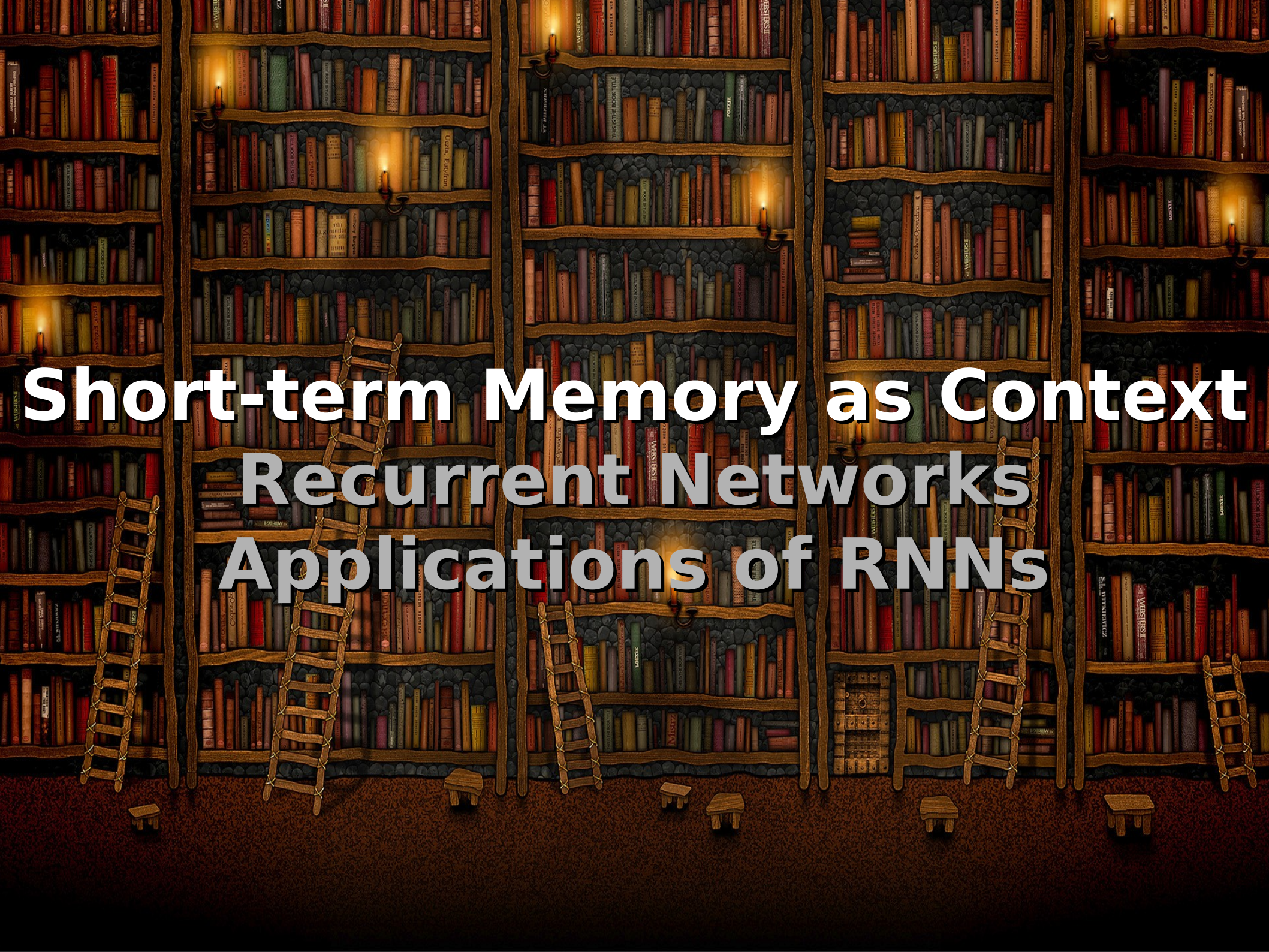
課號：Psy7277　　教室： 北 206
識別碼：227M9280　　時間： 五 234

Deep Neural Networks 的分類

CNN 通常處理影像資料；RNN 通常處理語言資料




本周介紹 RNN
！


Short-term Memory as Context
Recurrent Networks
Applications of RNNs

從知覺到動作

相同 / 類似的知覺不一定產生一樣的動作

紅	黑	綠	藍
黃	橙	黑	棕
紫	黃	藍	黃
綠	棕	紅	紫

數學上無解：

x= 紅 → y= 紅（叫意）

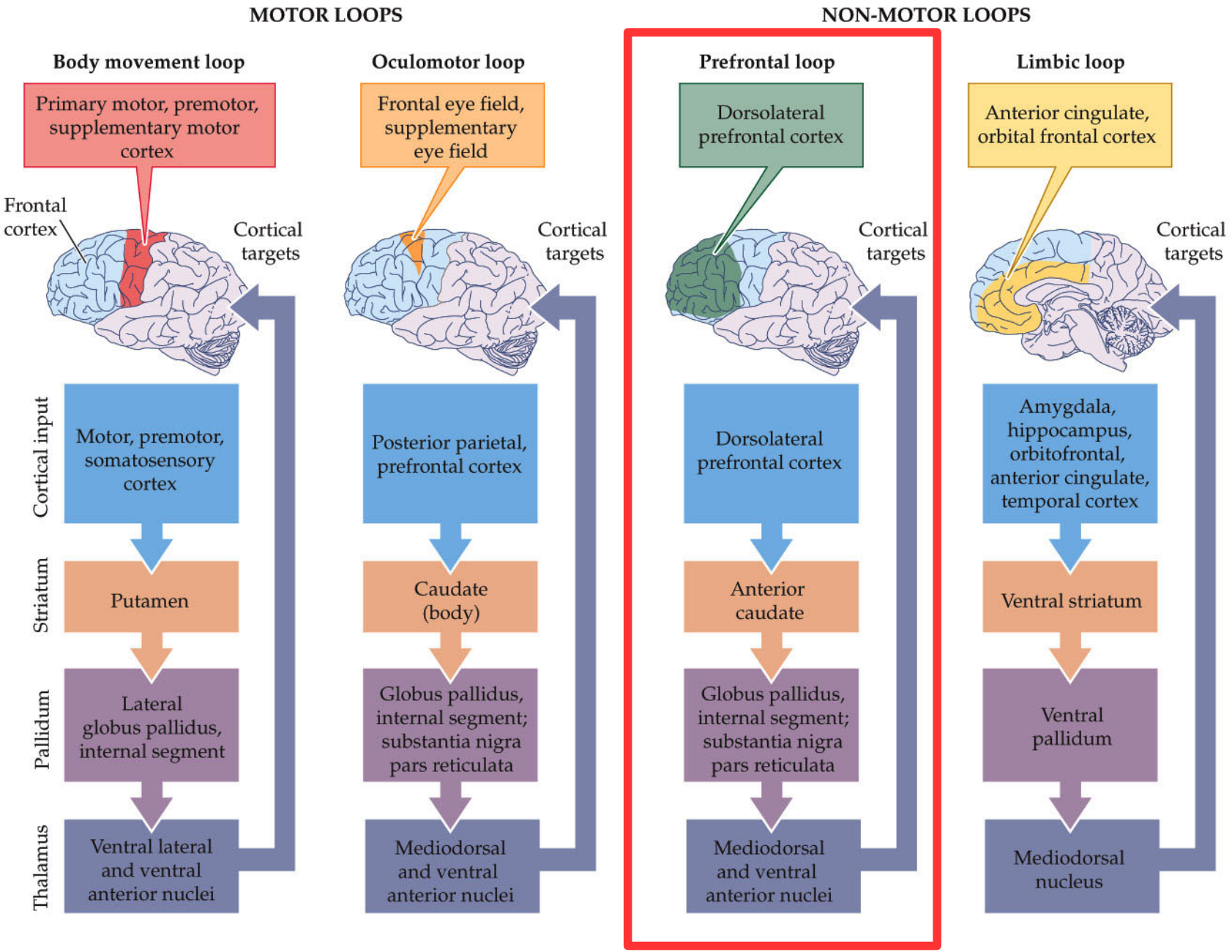
x= 紅 → y= 綠（叫色）

數學上有解：

x=(紅 , 叫意) → y= 紅

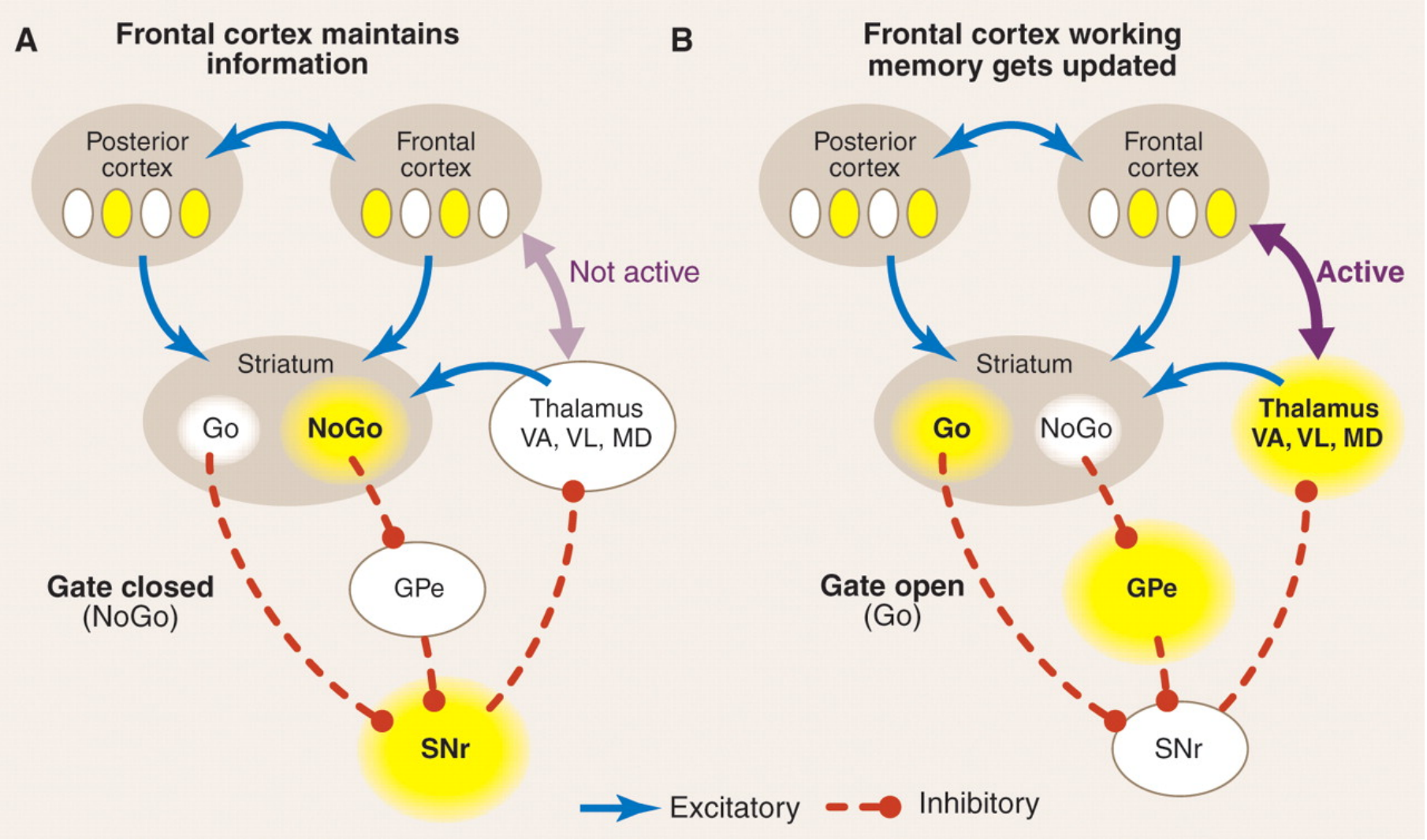
x=(紅 , 叫色) → y= 綠

Cortico-Basal Ganglia Loops



BG-modulated PFC

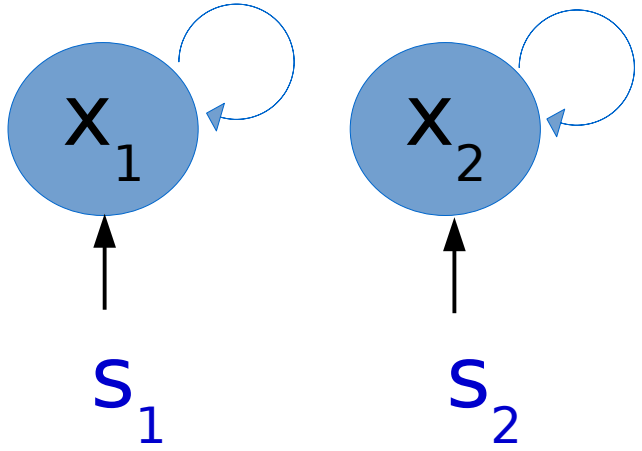
= Prefrontal cortex Basal ganglia Working Memory




Short-term Memory as Context
Recurrent Networks
Applications of RNNs

Self-Recurrent Excitations

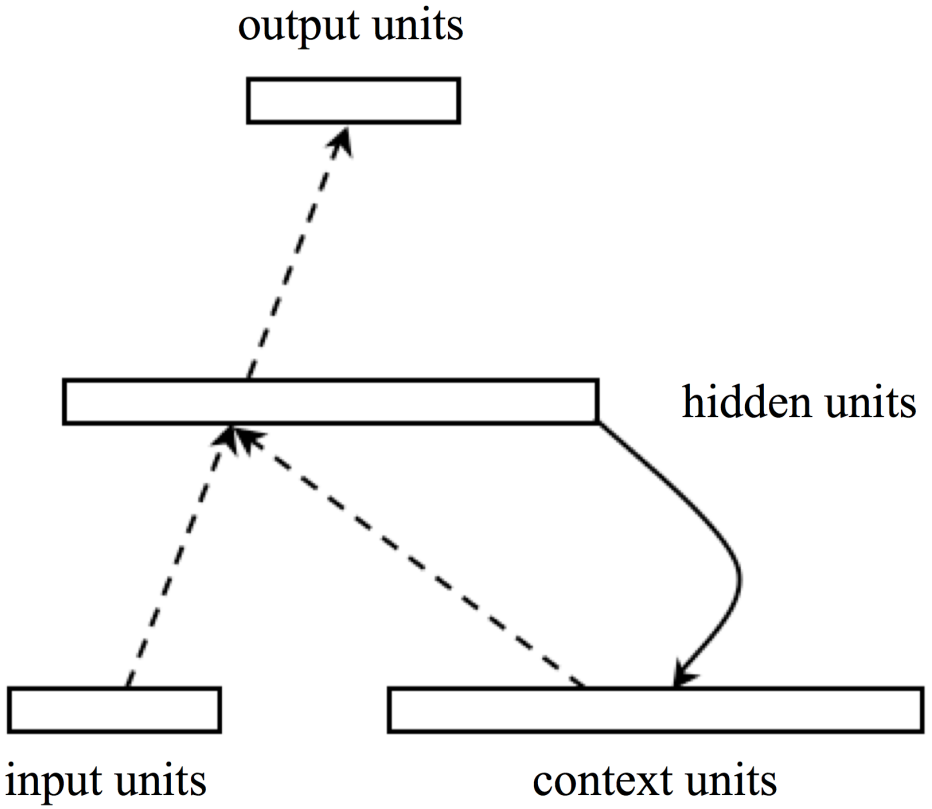
自我連結的刺激可維持反應卻無法維持刺激對比

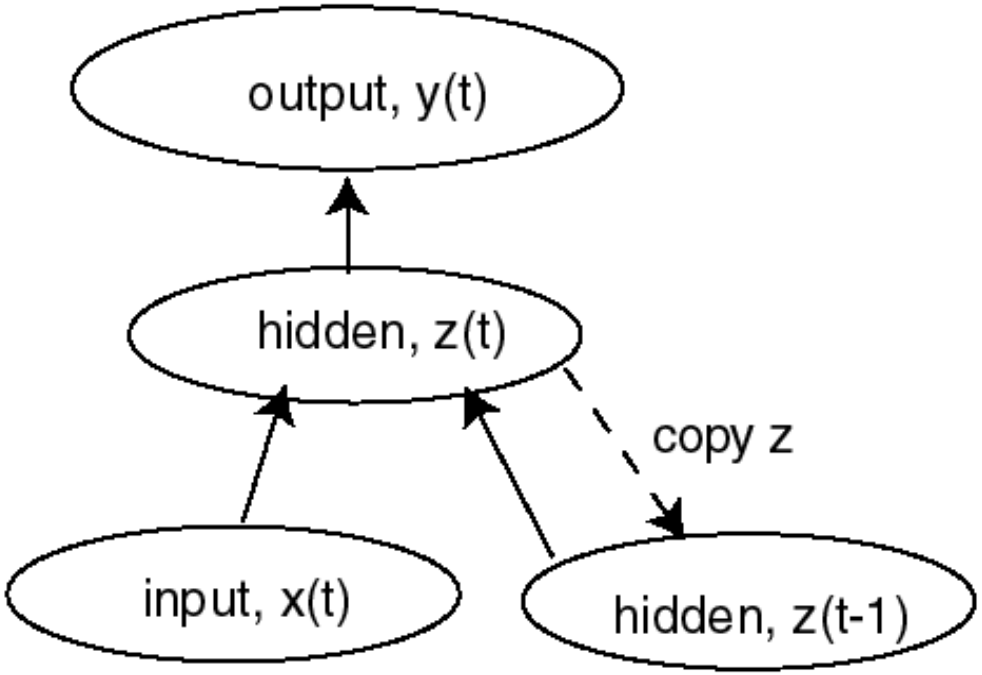



```
x=[0,0]; dt=0.1
for t in arange(0,10,dt):
    s=[1,10] if t<1 else [0,0]
    x[0]=x[0]+dt*(-0.1*x[0]+(1-x[0])*(s[0]+x[0]))
    x[1]=x[1]+dt*(-0.1*x[1]+(1-x[1])*(s[1]+x[1]))
    clf(); plot([1,2],x,'-o')
    ylim([0,1]); title('t='+str(t))
    display(gcf()); clear_output(wait=True)
```

Simple RNN (1/2)

by Elman (1990)

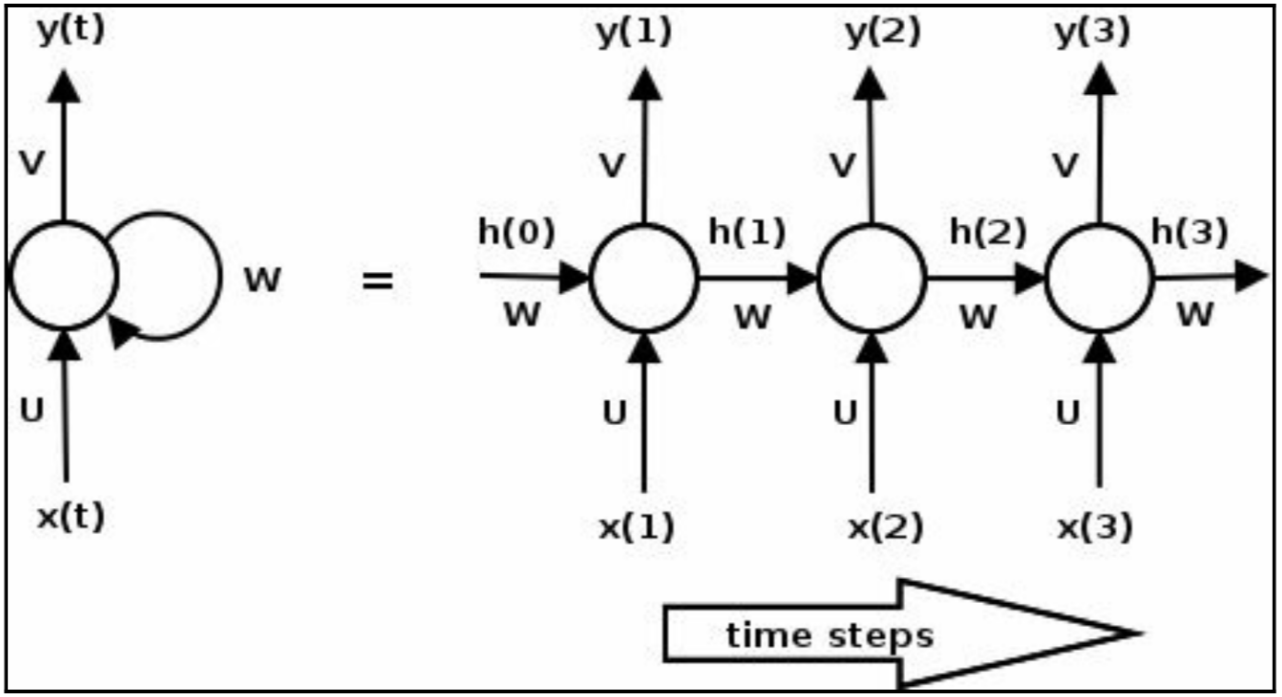




to process sequential information

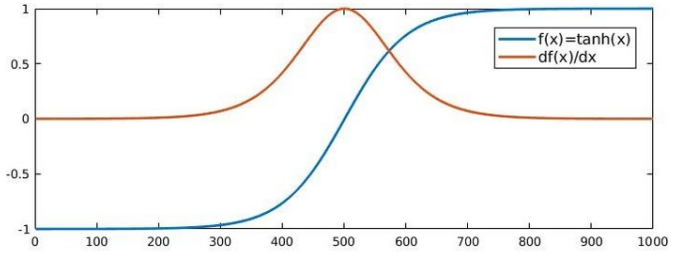
Simple RNN (2/2)

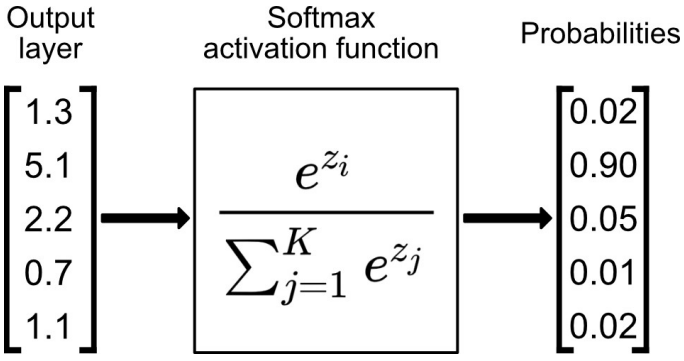
effectively has VERY deep layers



$$h_t = \tanh(Wh_{t-1} + UX_t)$$

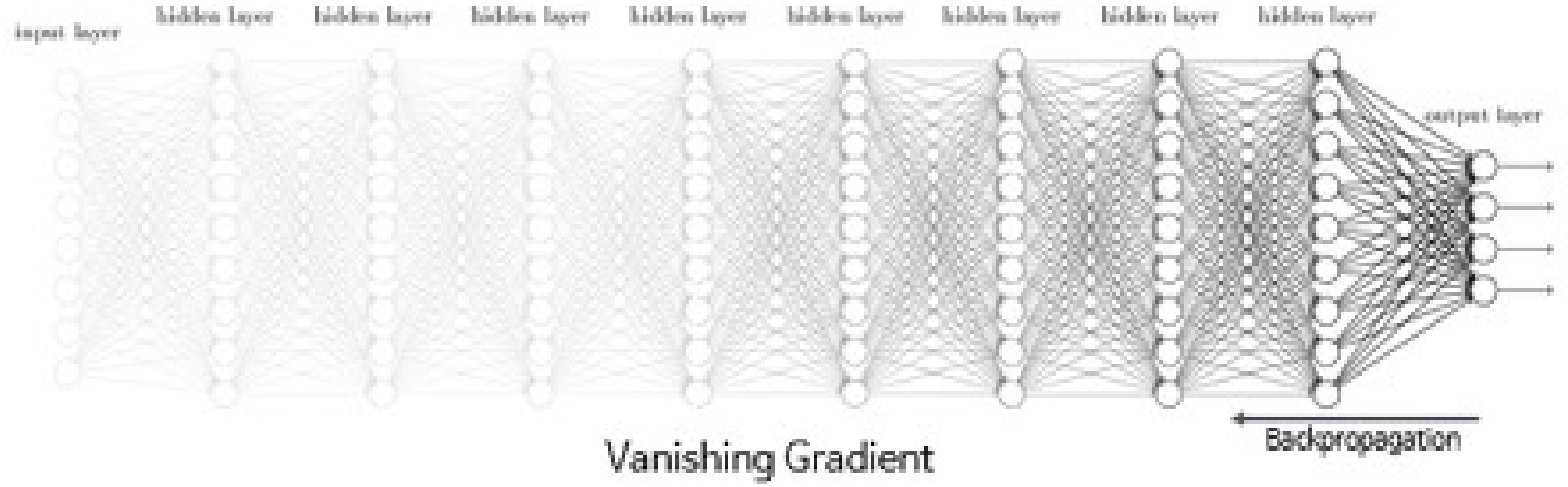
$$y_t = \text{softmax}(Vh_t)$$

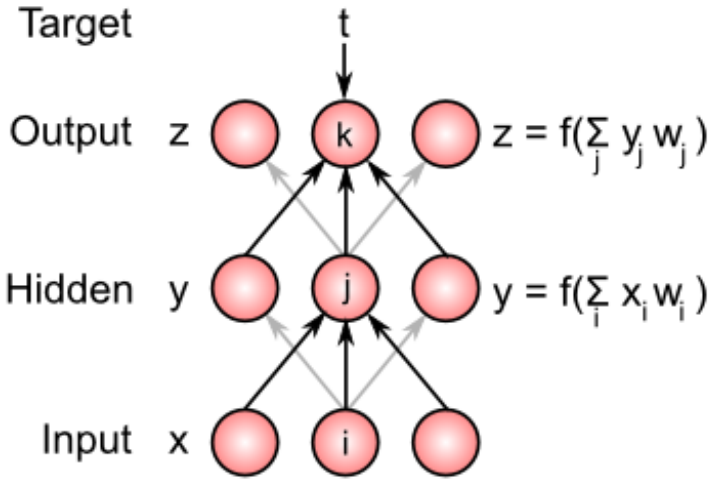




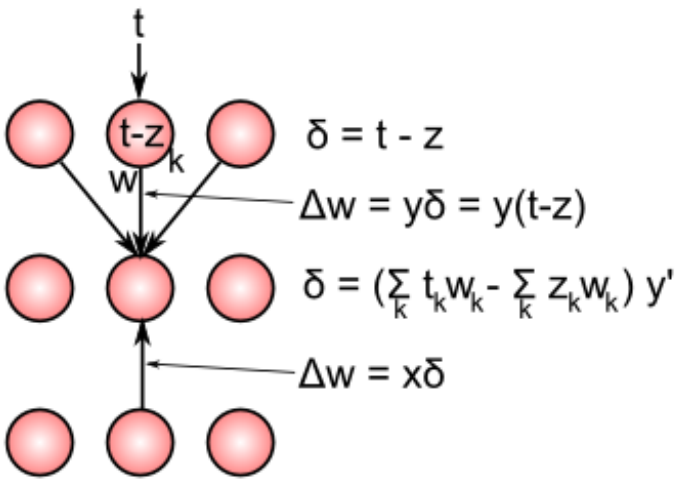
Temporal Backpropagation

suffers from vanishing & exploding gradient problem

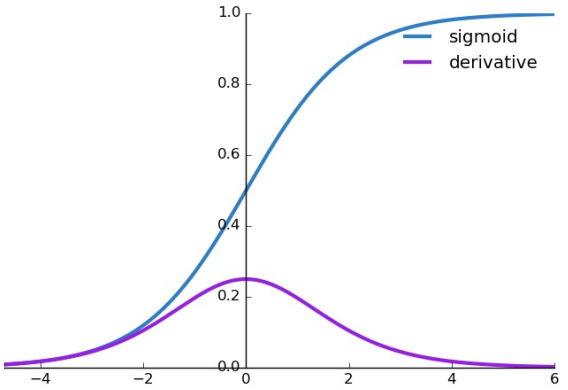




a) Feedforward Activation

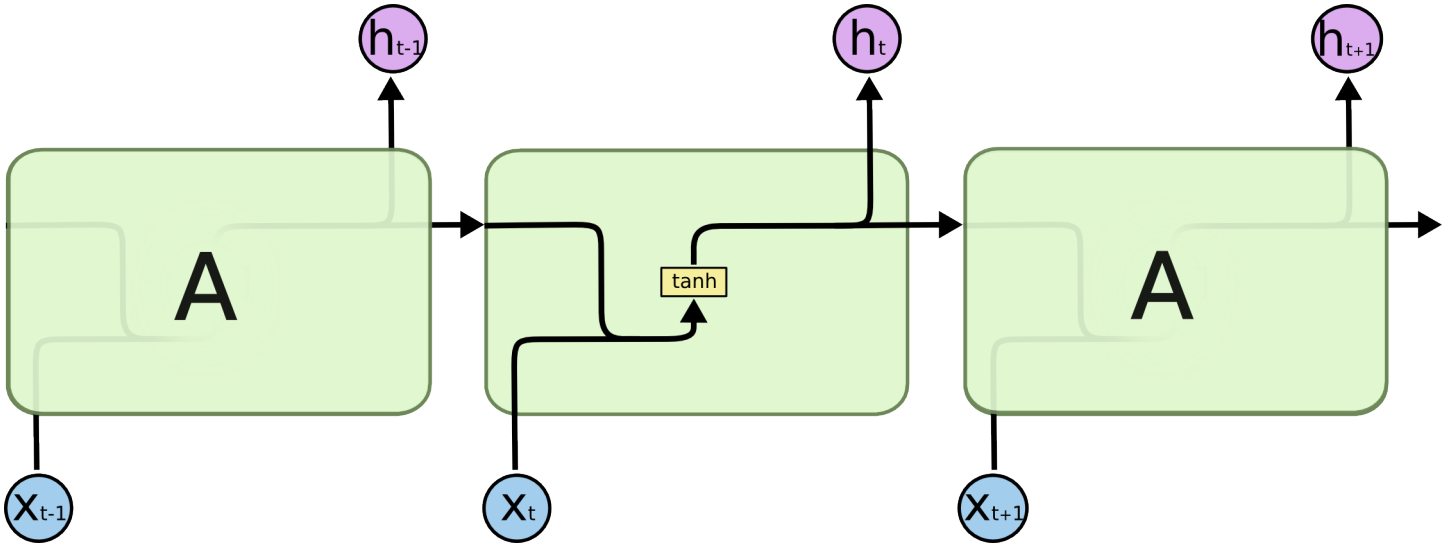


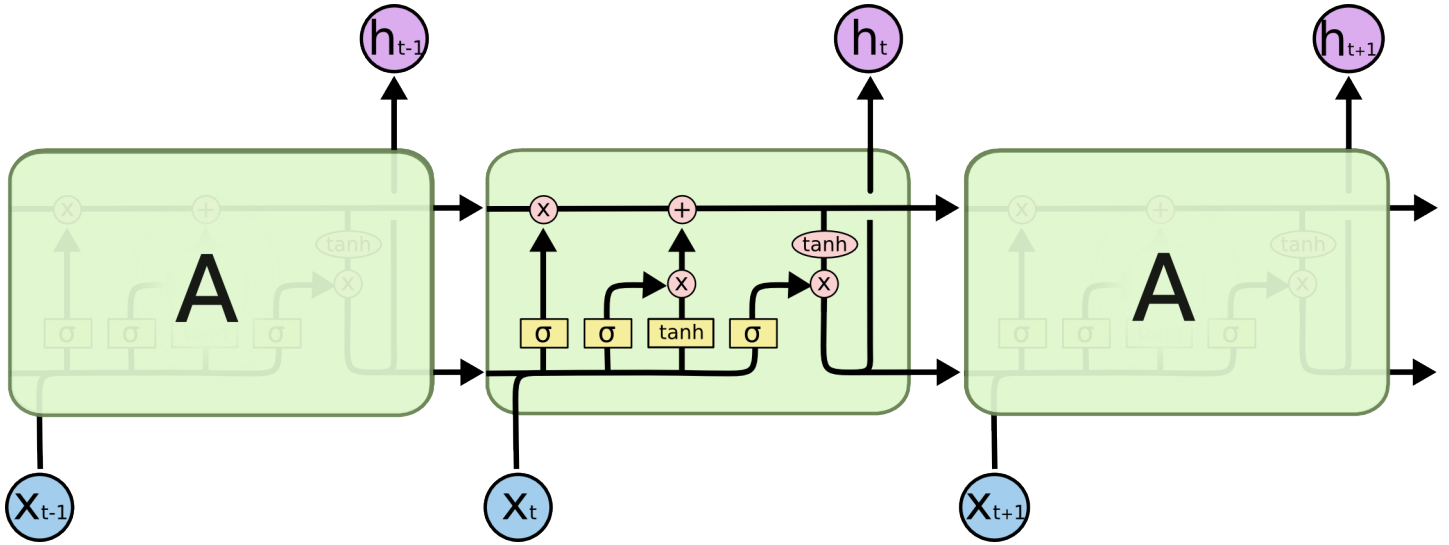
b) Error Backpropagation



Long Short-term Memory (1/2)

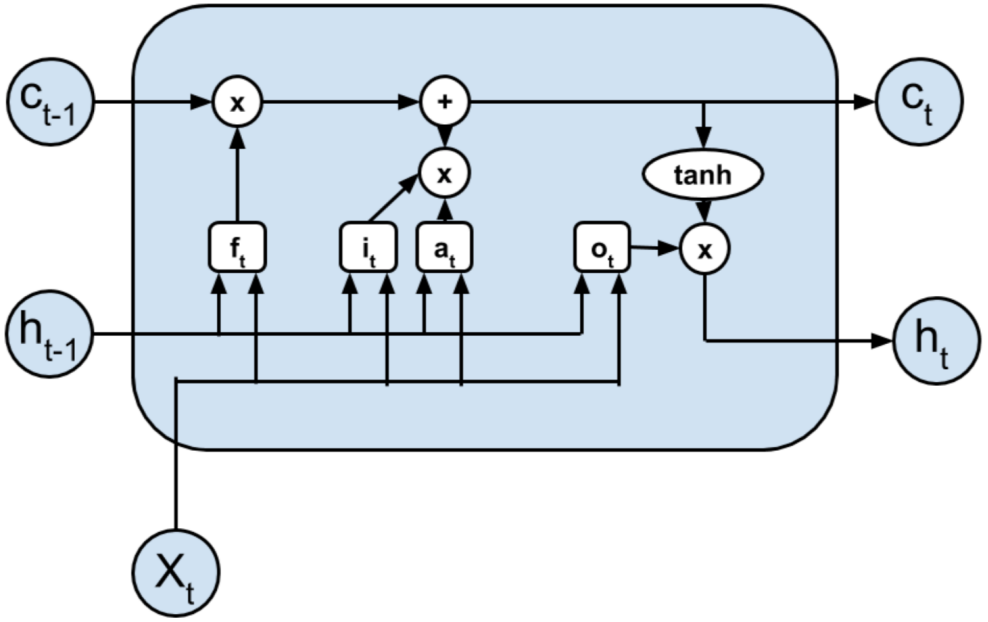
introduces input, forget, output gates





Long Short-term Memory (2/2)

Example: ignore '\n' & reset context for '.'



$$f_t = \sigma\left(W_f h_{t-1} + U_f x_t + b_f\right)$$

$$i_t = \sigma\left(W_i h_{t-1} + U_i x_t + b_i\right)$$

$$a_t = \tanh\left(W_c h_{t-1} + U_c x_t + b_c\right)$$

$$o_t = \sigma\left(W_o h_{t-1} + U_o x_t + b_o\right)$$

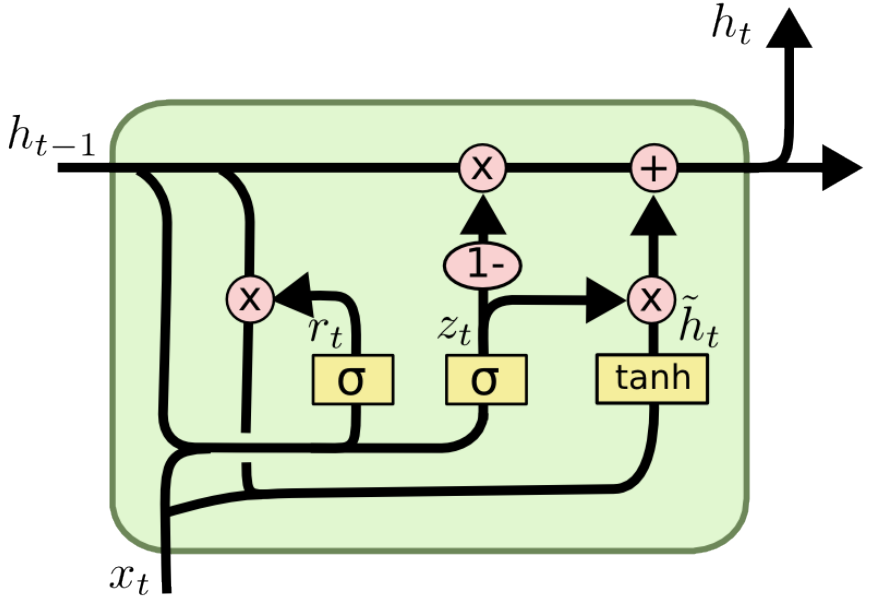
$$c_t = f_t * c_{t-1} + i_t * a_t$$

$$h_t = o_t * \tanh\left(c_t\right)$$

$(f_t, i_t)=(1,0)$ to maintain perfect memory

Gated Recurrent Unit

GRU is a simplified LSTM without an output gate



$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t]\right)$$

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t]\right)$$

$$\tilde{h}_t = \tanh\left(W \cdot [r_t * h_{t-1}, x_t]\right)$$

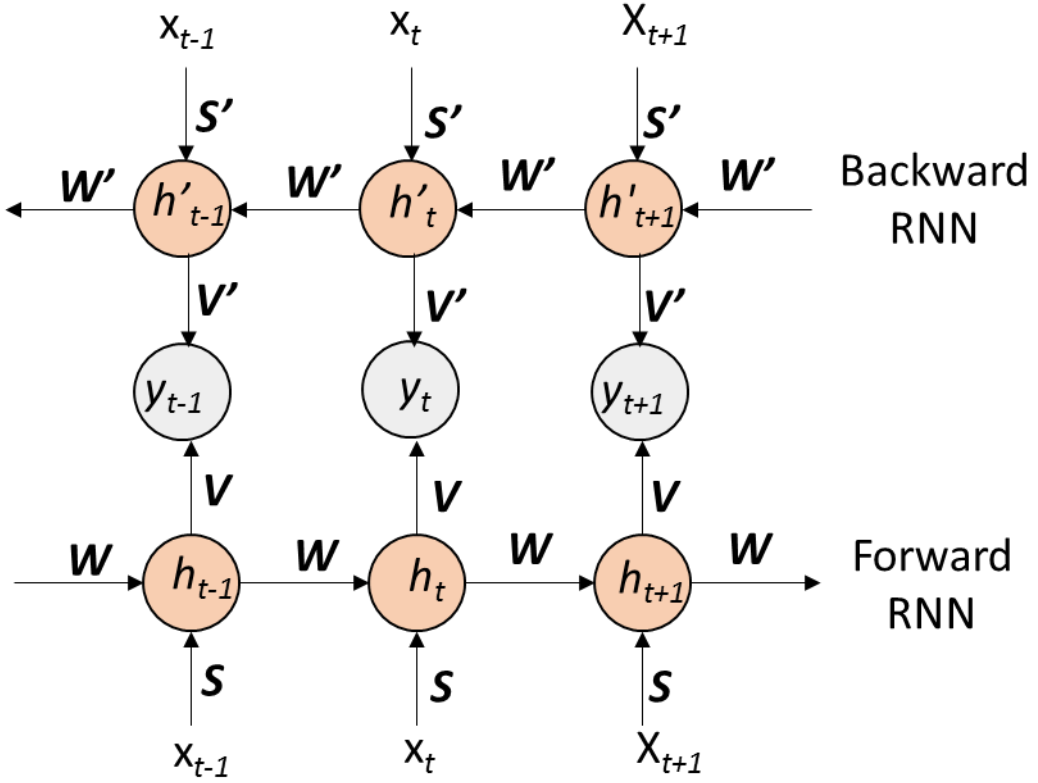
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

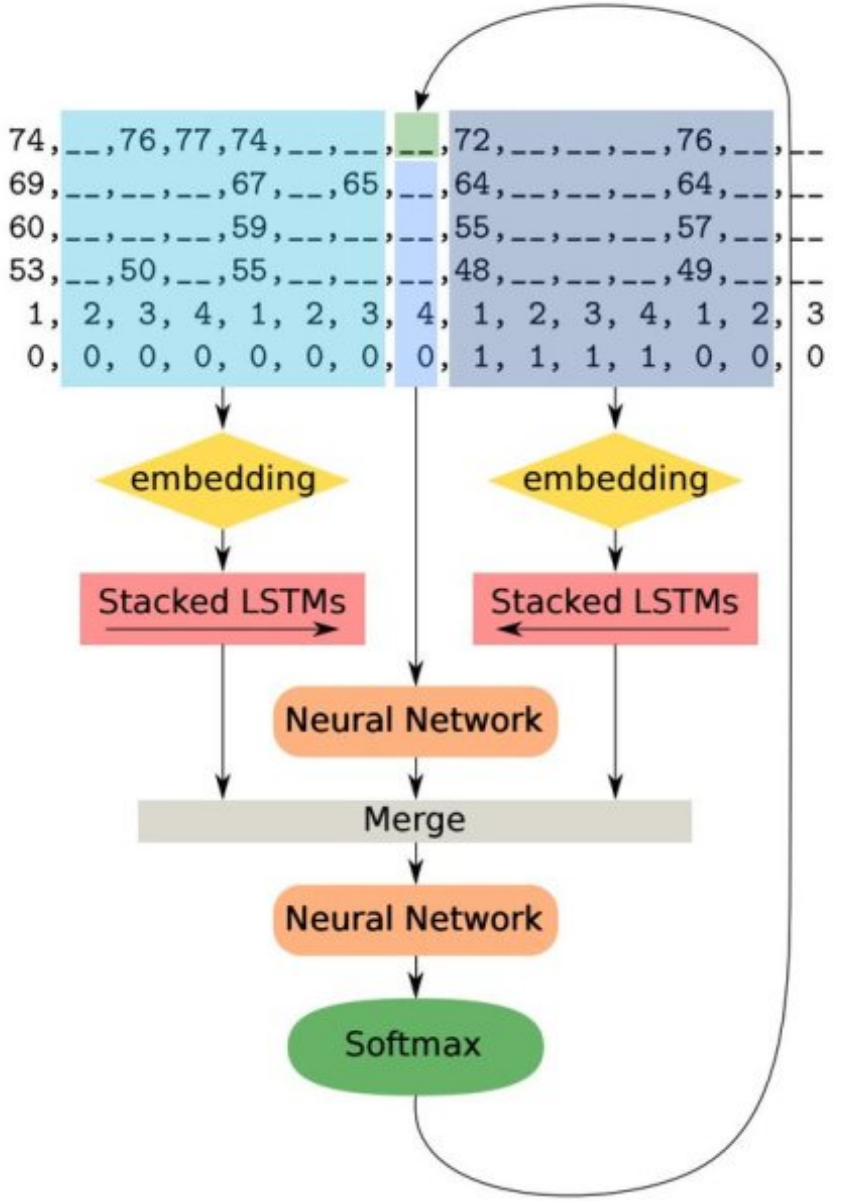
memory input

r_t = reset/forget gate ; z_t = update/input gate

Bidirectional RNN

DeepBach uses a similar idea





Attention replacing RNN (1/3)

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com

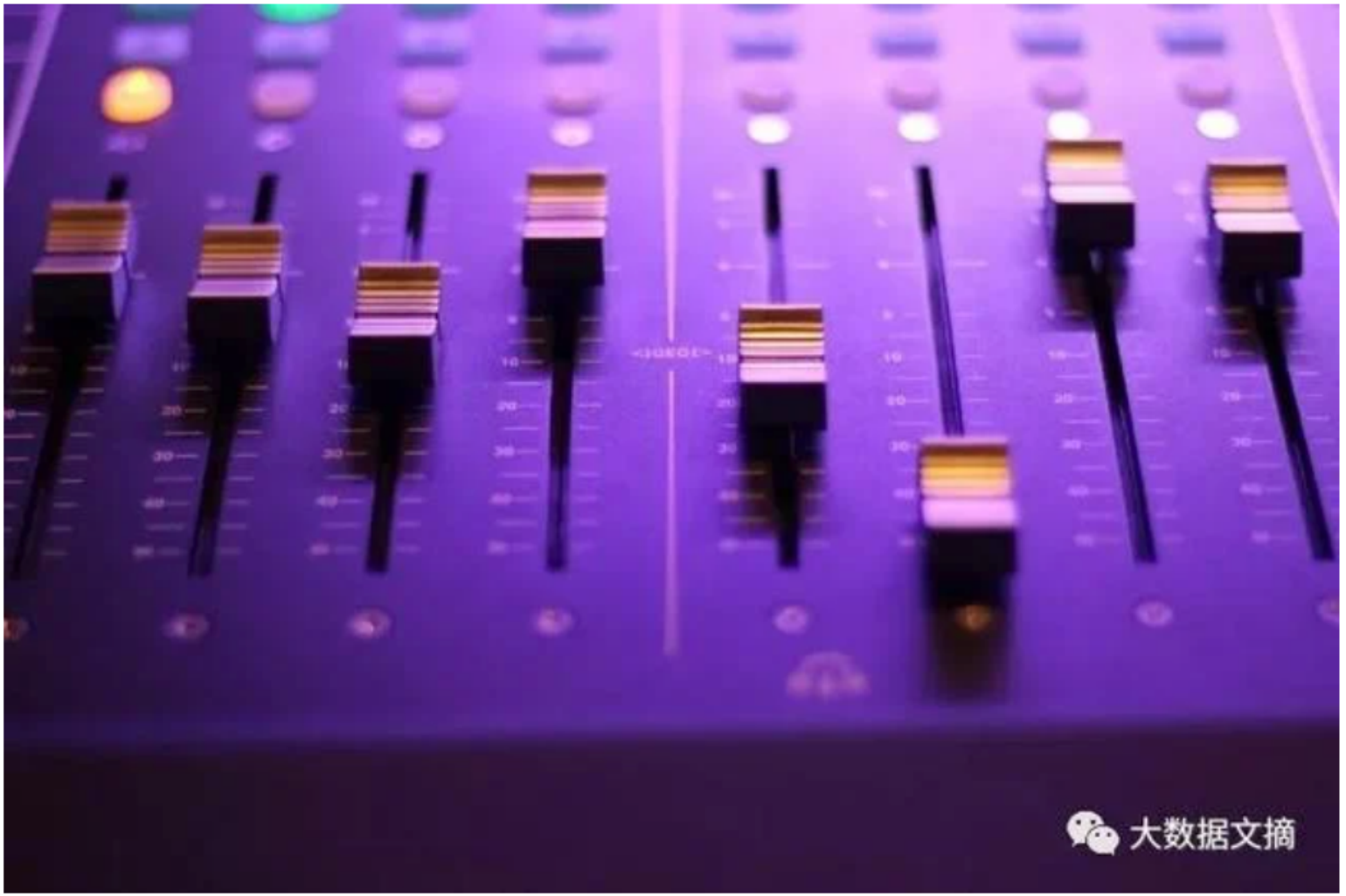
Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.0 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature.

Attention replacing RNN (2/3)

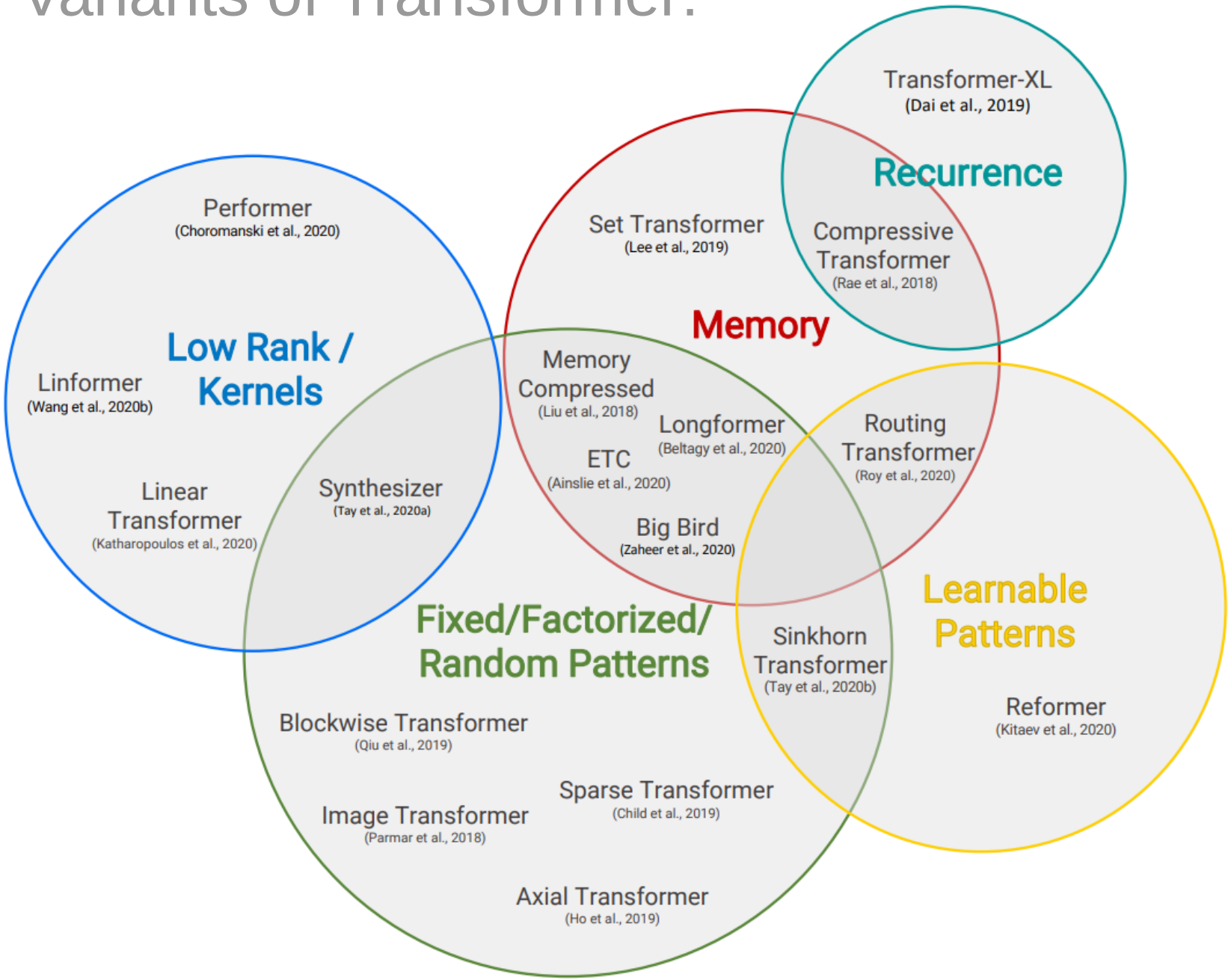
别再用RNN和LSTM了！注意力模型才是王道

2018-05-02 *由 大數據文摘 發表于科技*

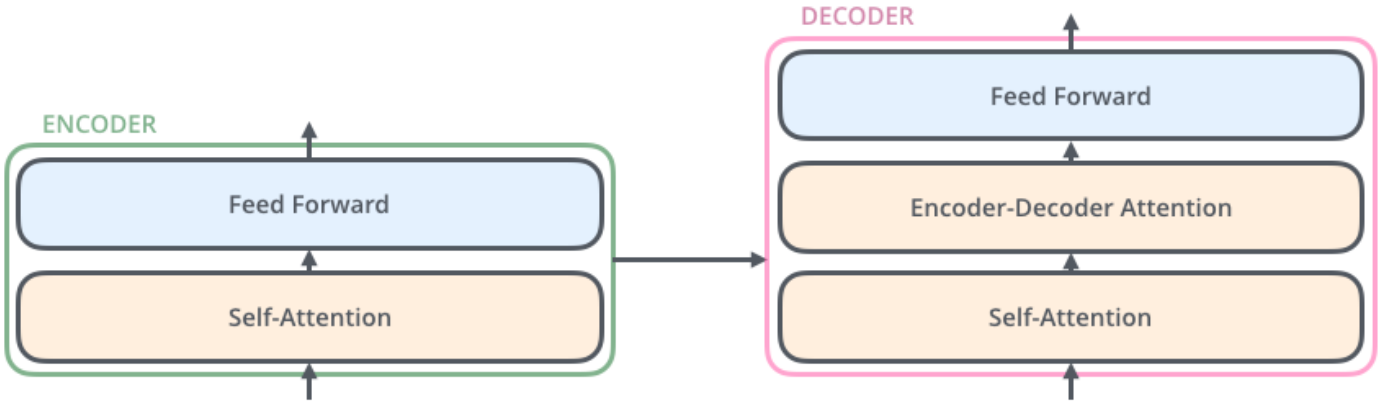


Attention replacing RNN (3/3)

Variants of Transformer:



Attention in Transformer



Transformer

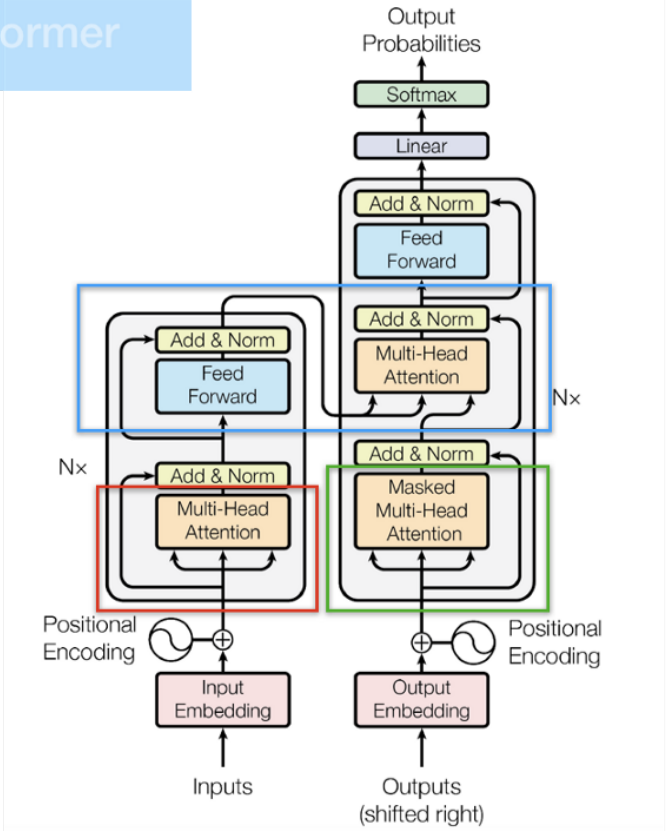


Figure 1: The Transformer - model architecture.

encoder self attention

1. Multi-head Attention
2. **Q**uery=**K**ey=**V**alue

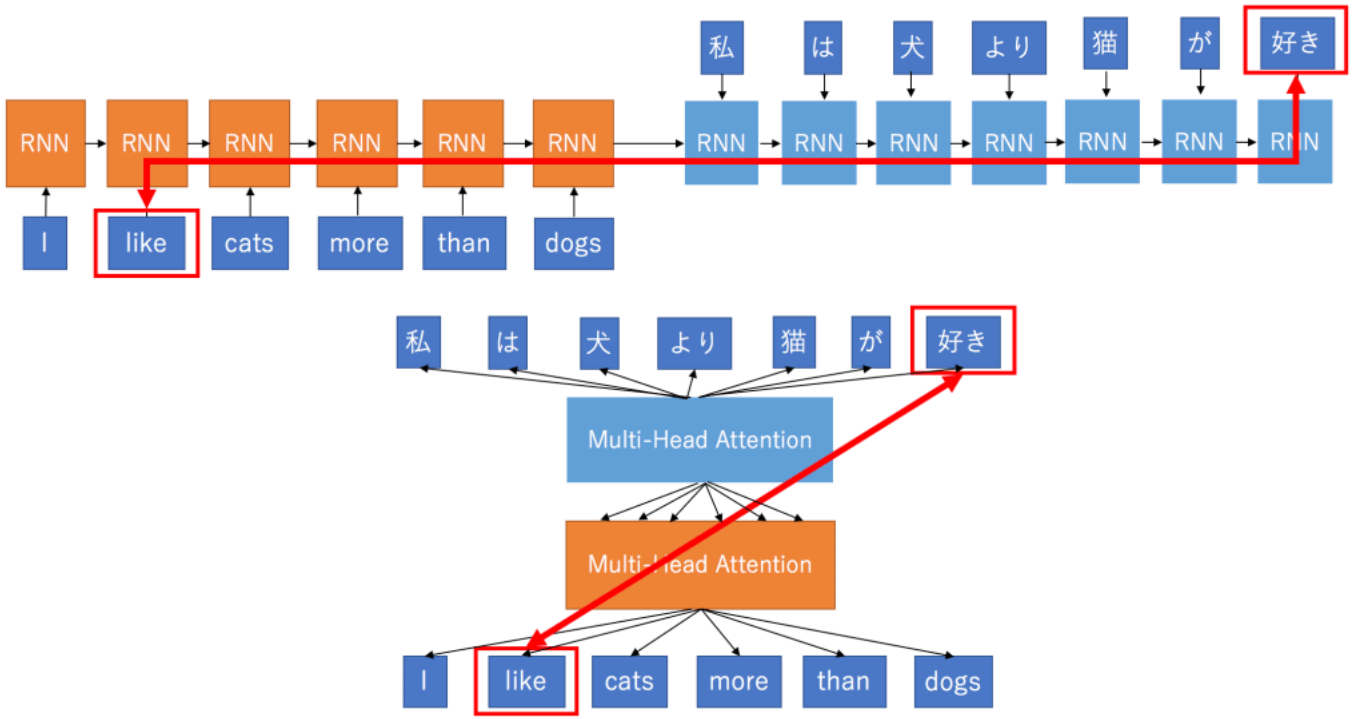
decoder self attention

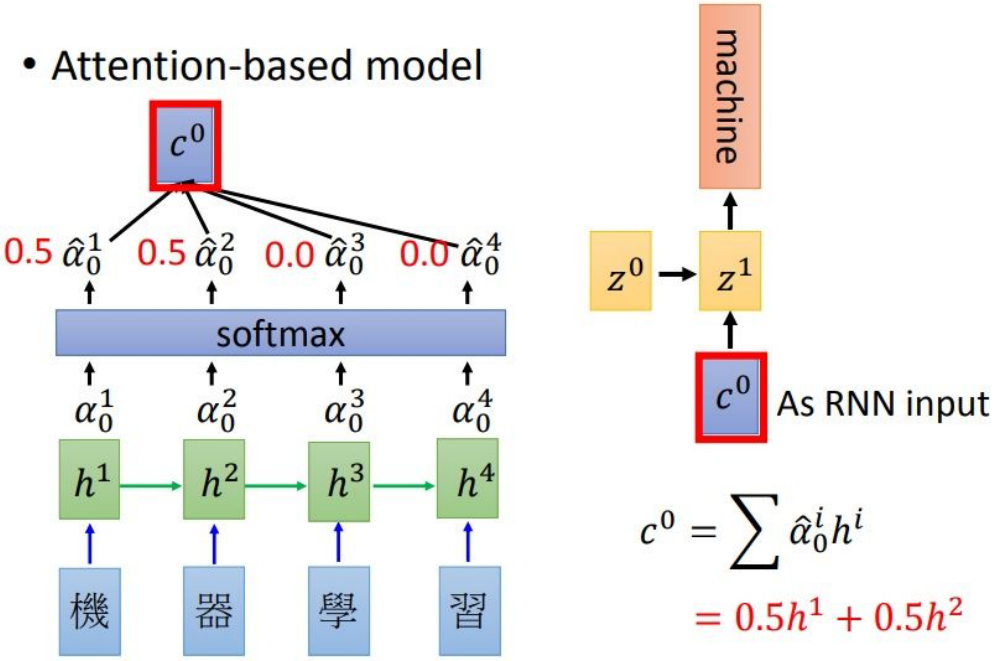
1. **Masked** Multi-head Attention
2. **Q**uery=**K**ey=**V**alue

encoder-decoder attention

1. Multi-head Attention
2. Encoder Self attention=**K**ey=**V**alue
3. Decoder Self attention=**Query**

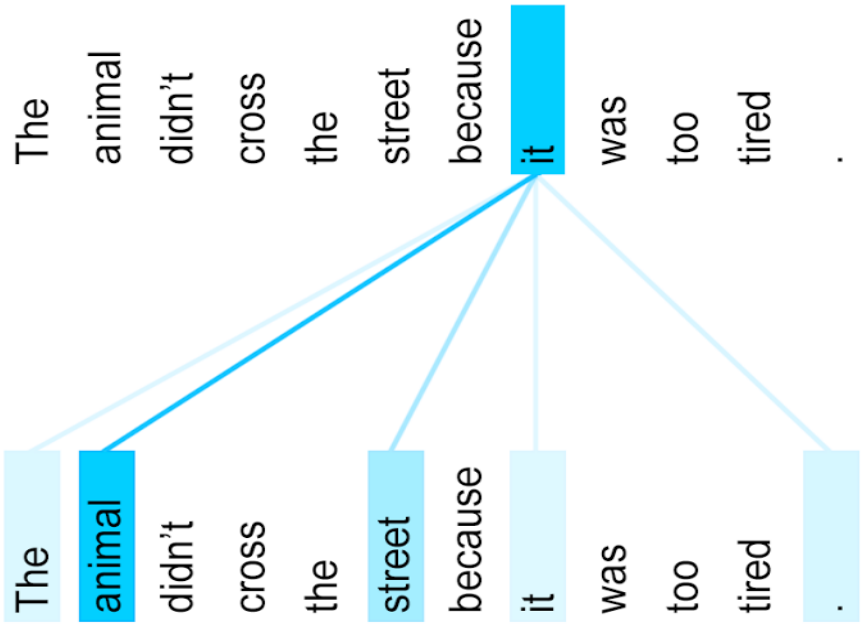
Encoder-Decoder Attention

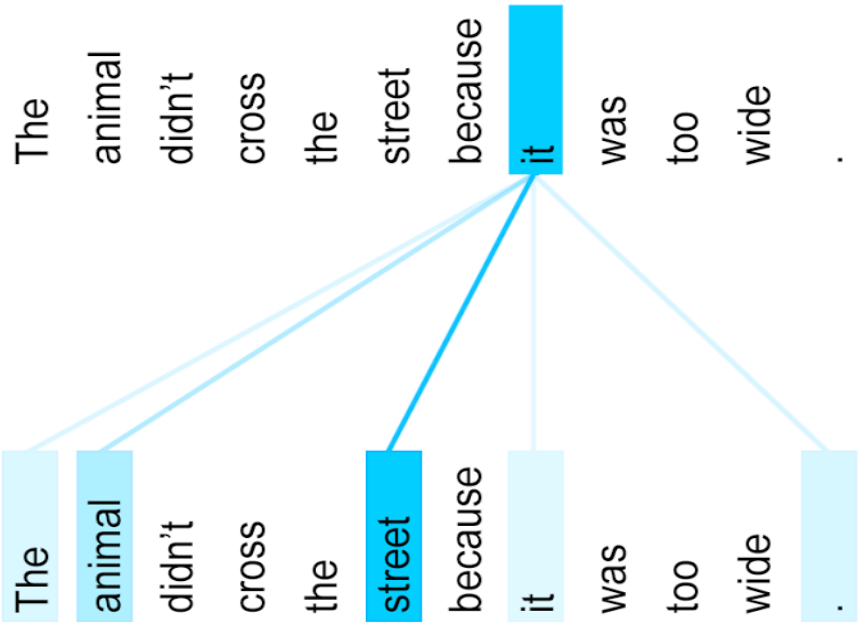




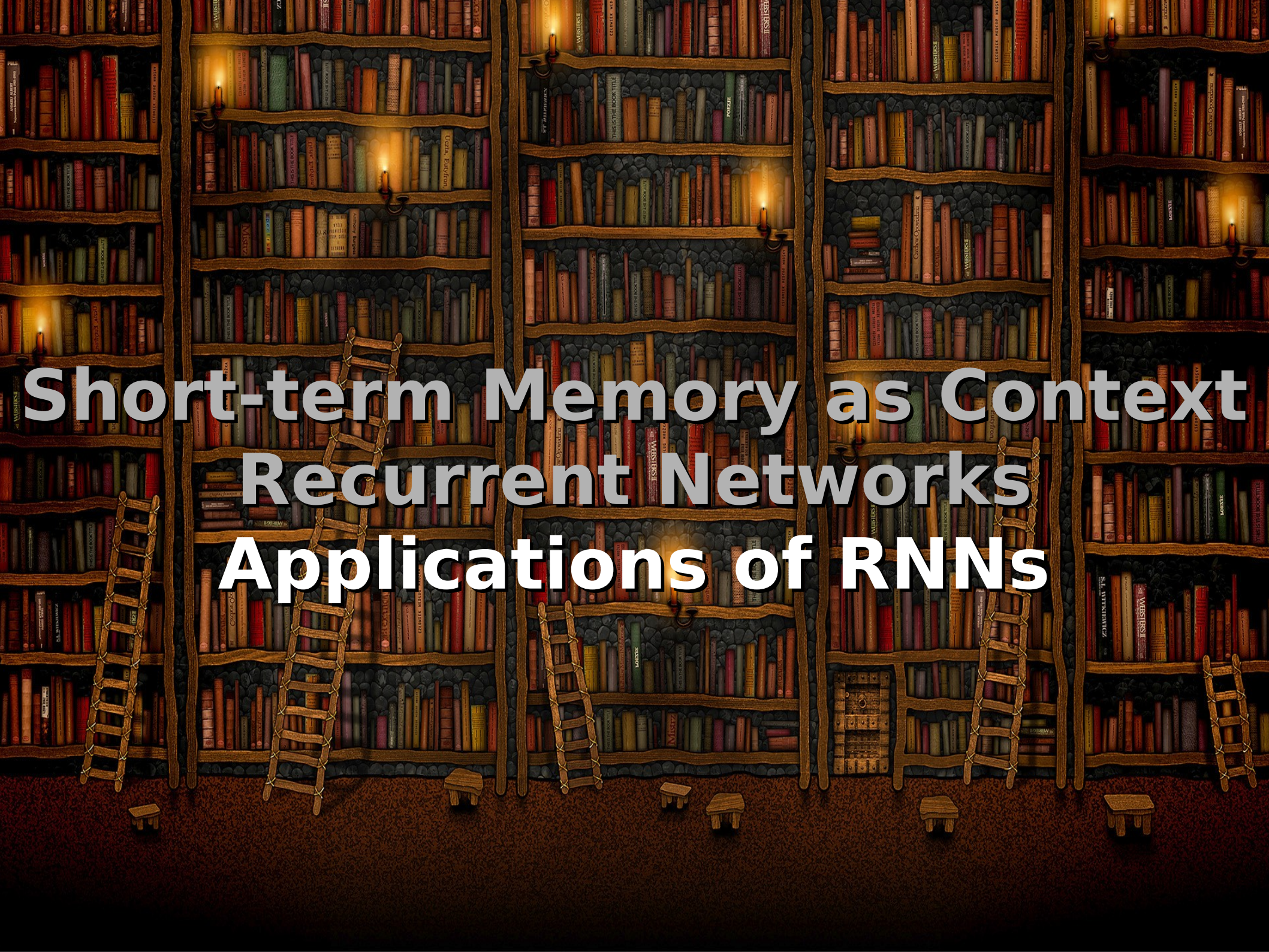
Encode/Decoder Self-Attention

Example: What's the correct understanding of "it"?



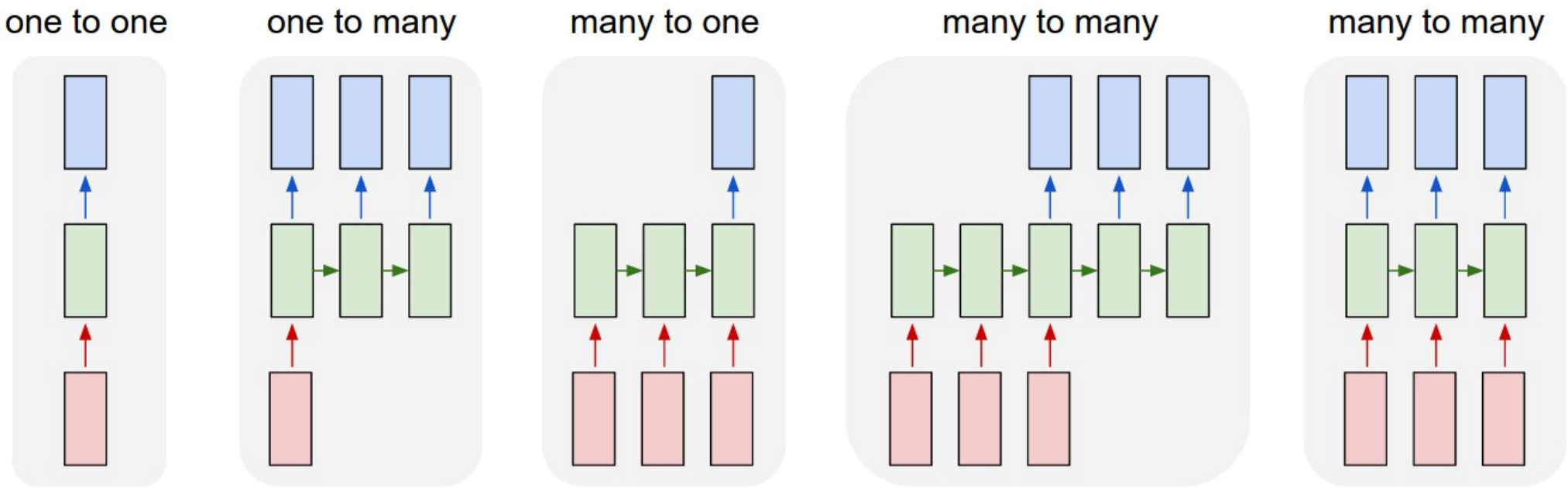


Different understanding of "it" = Different vector of "it"


Short-term Memory as Context
Recurrent Networks
Applications of RNNs

Recurrent Neural Networks

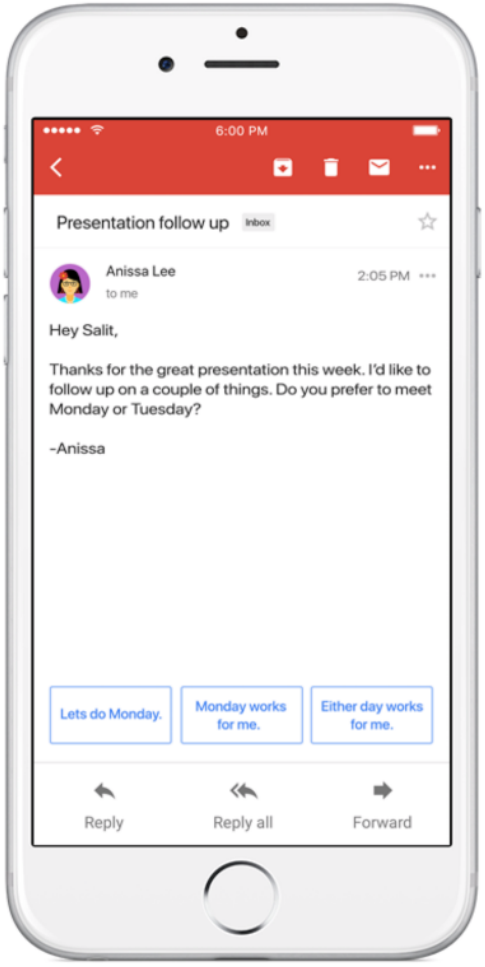
take many forms



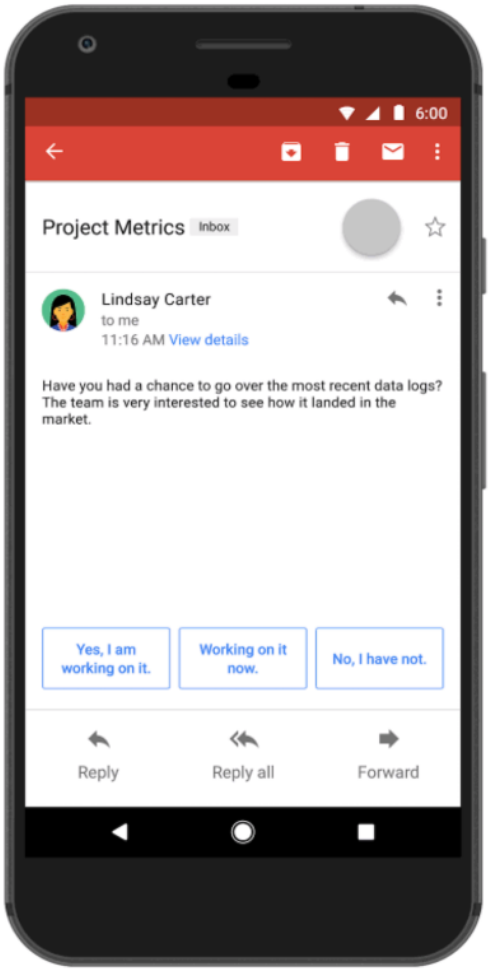
for image classification, image captioning, sentiment classification, language translation, speech recognition, etc.

Google SmartReply

Guessing what you want to reply

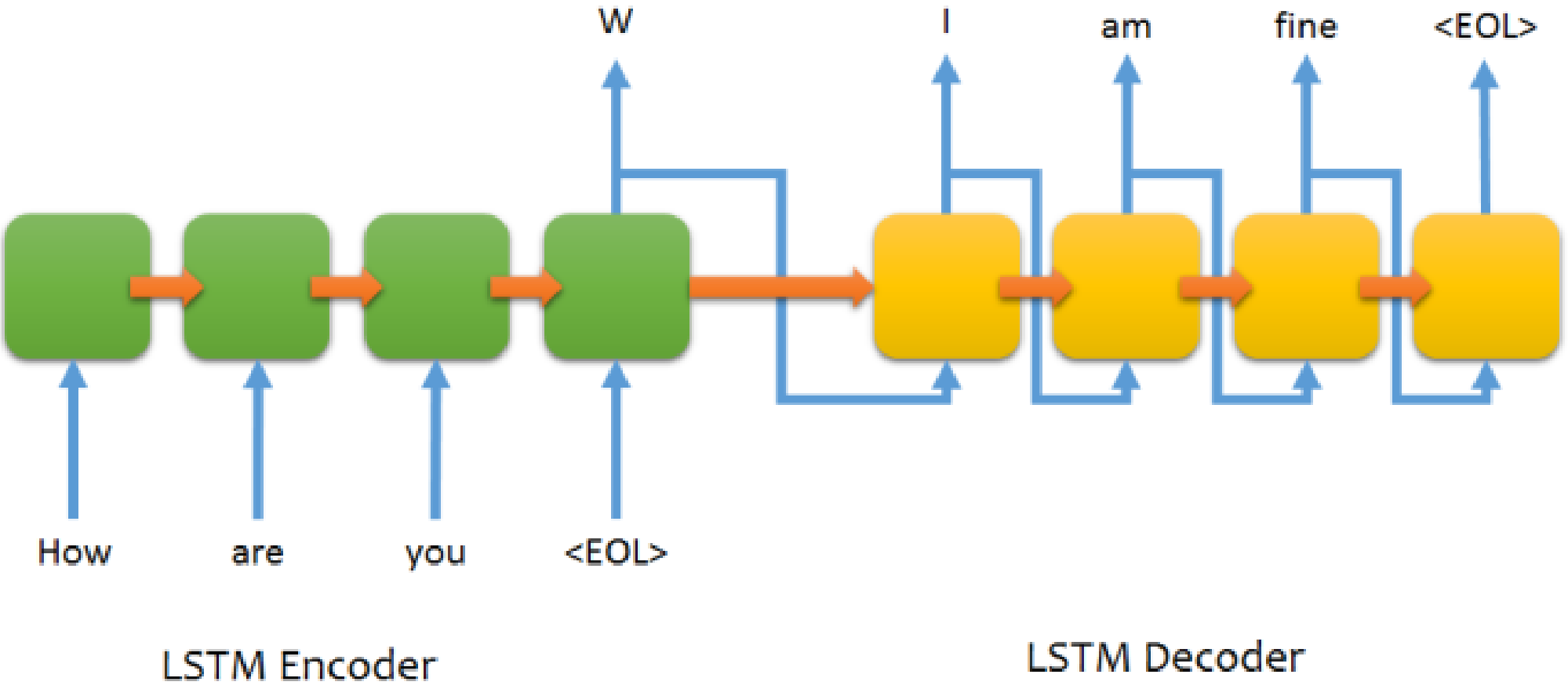


Smart Reply in Gmail



Chatbots (using Seq2Seq)

Guessing what you want to know

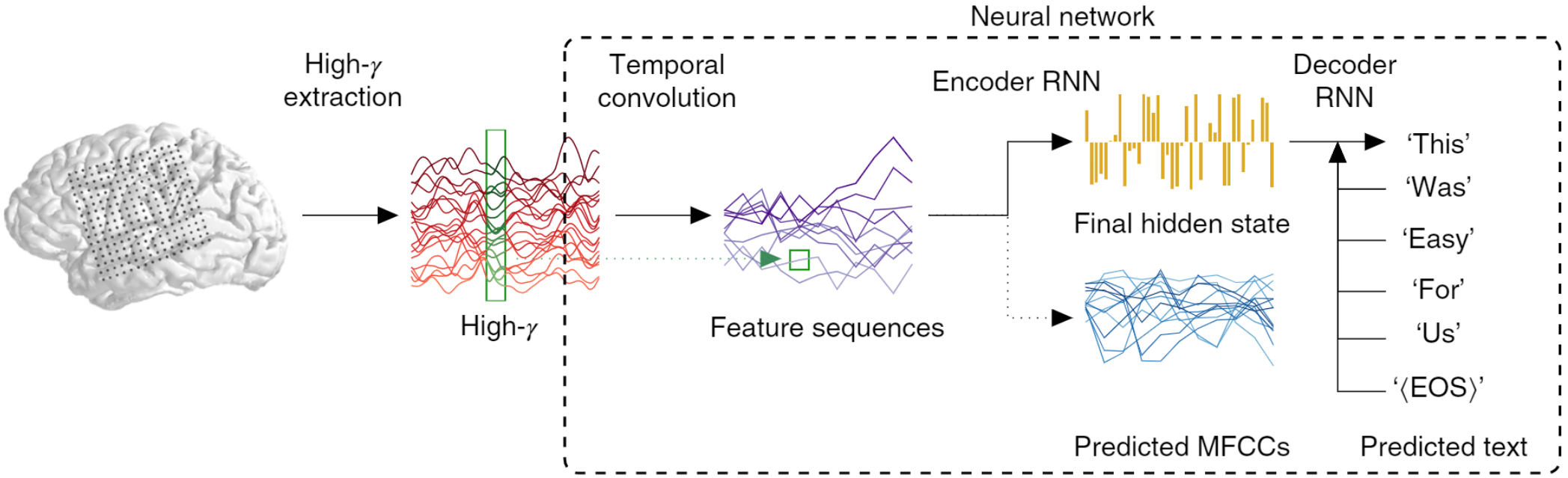


Data Modeling 範例

利用循環神經網路 (RNN) 來做 Brain Decoding

Machine translation of cortical activity to text with an encoder–decoder framework

Joseph G. Makin [1,2] ✉, David A. Moses[1,2] and Edward F. Chang [1,2] ✉



Generating Scripts for Movies

Here is the Chinese version of Sunspring

Speech Synthesis

The reverse problem of Speech Recognition

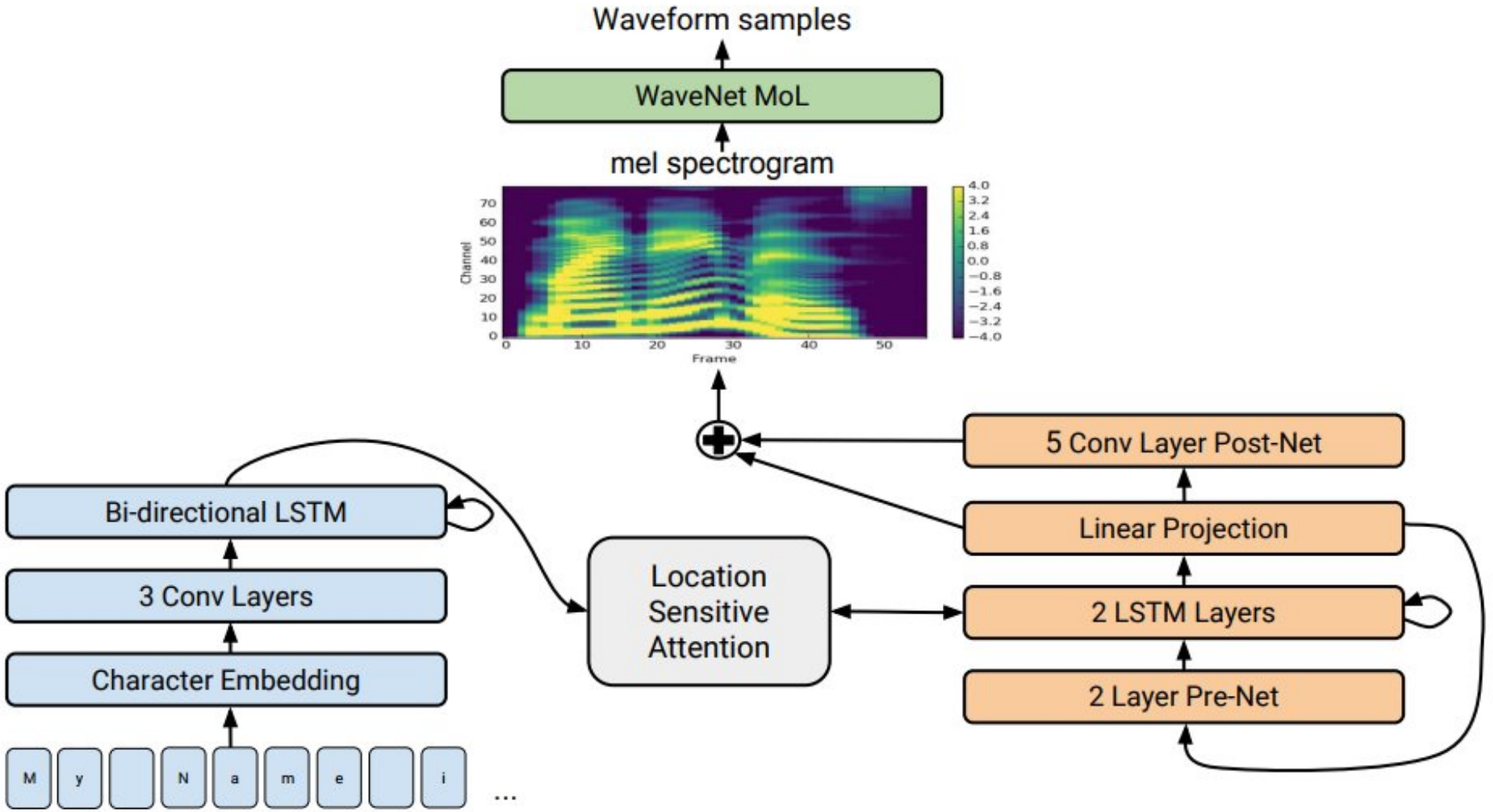
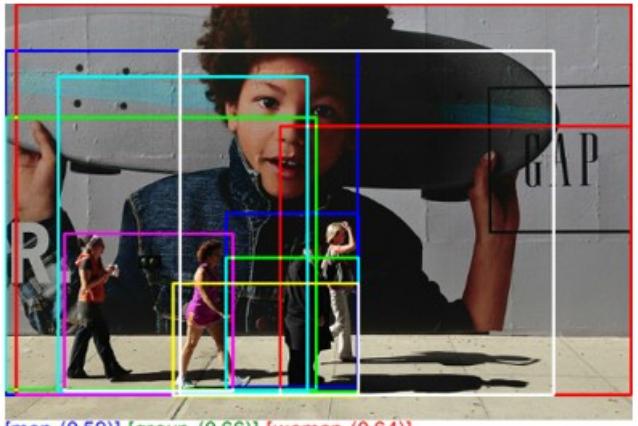
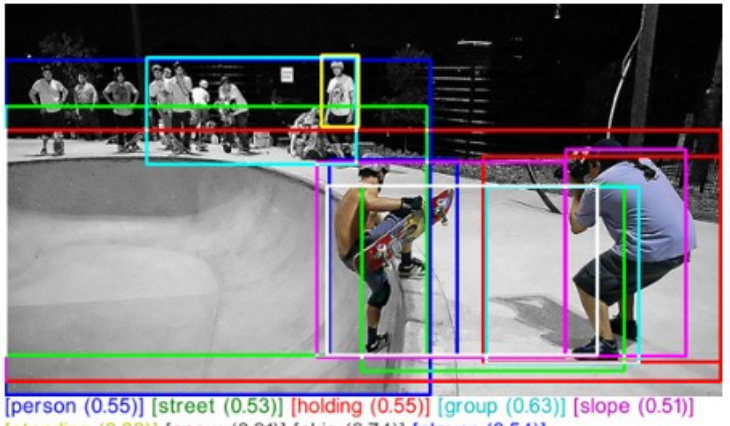
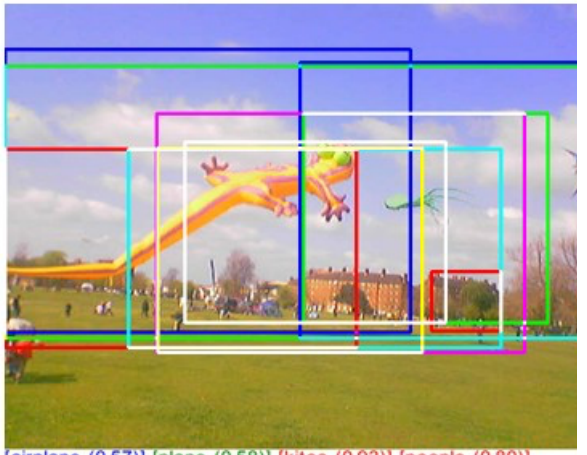


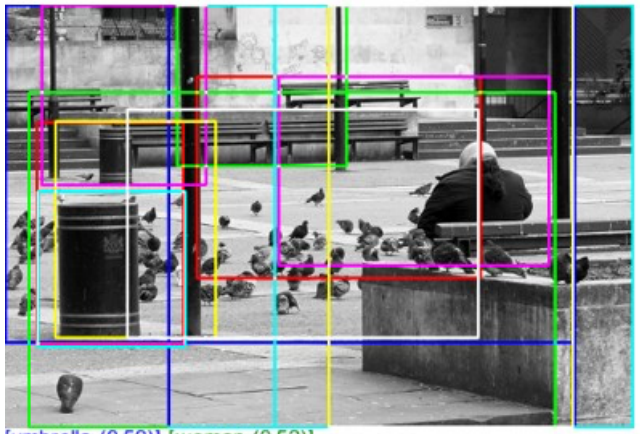
Image Captioning

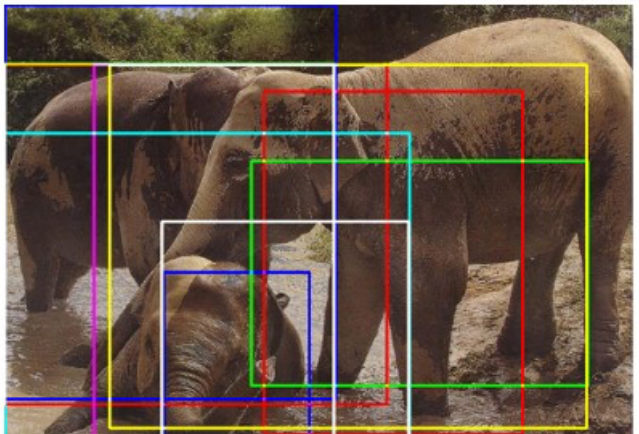
Image to text

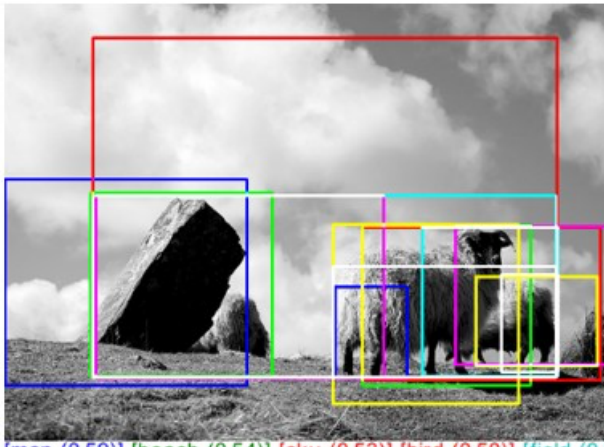






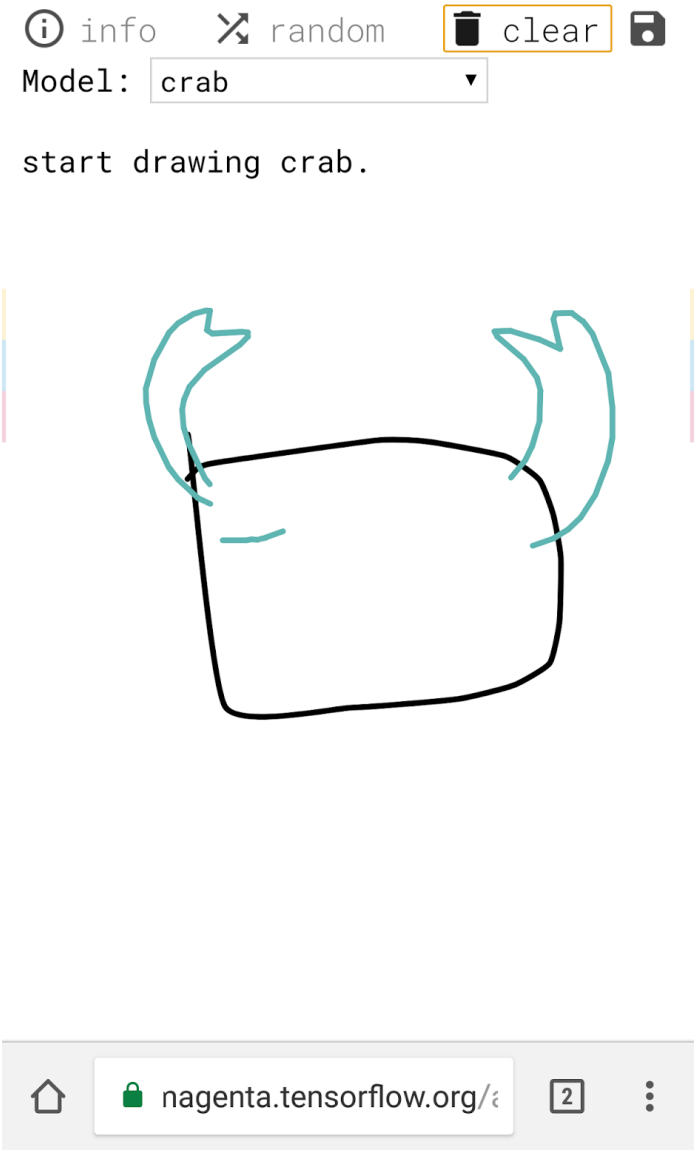






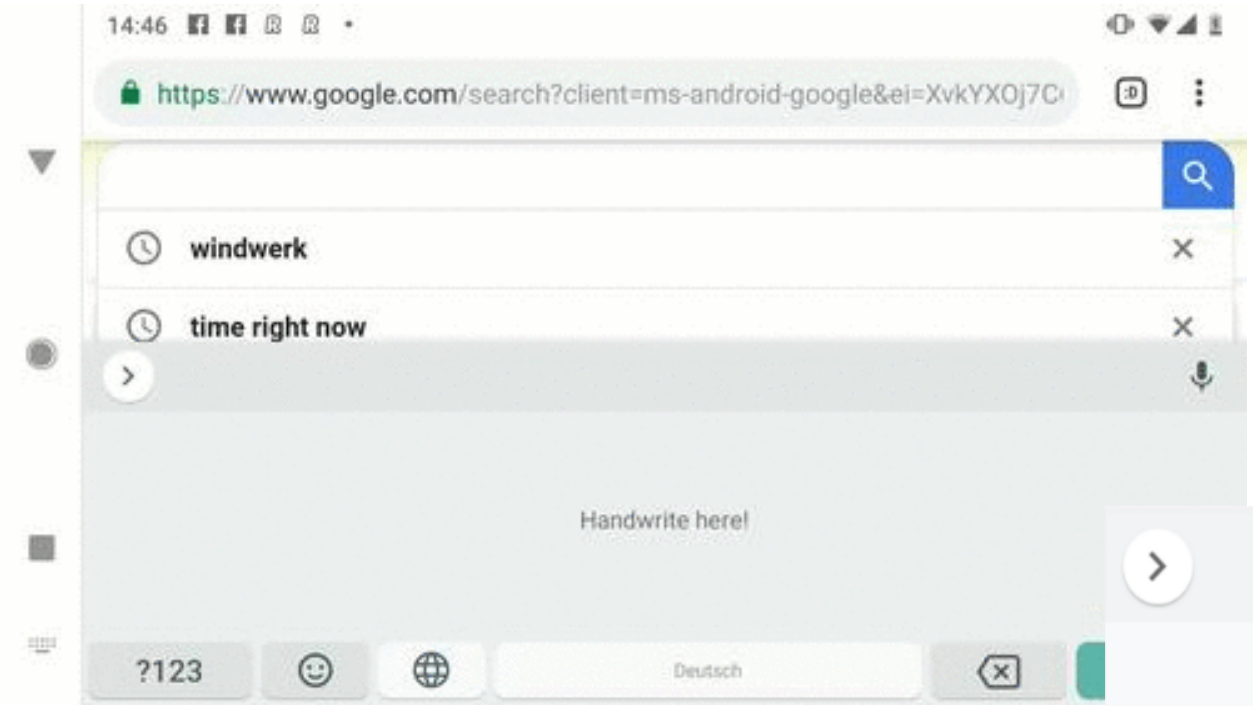
Sketch-RNN

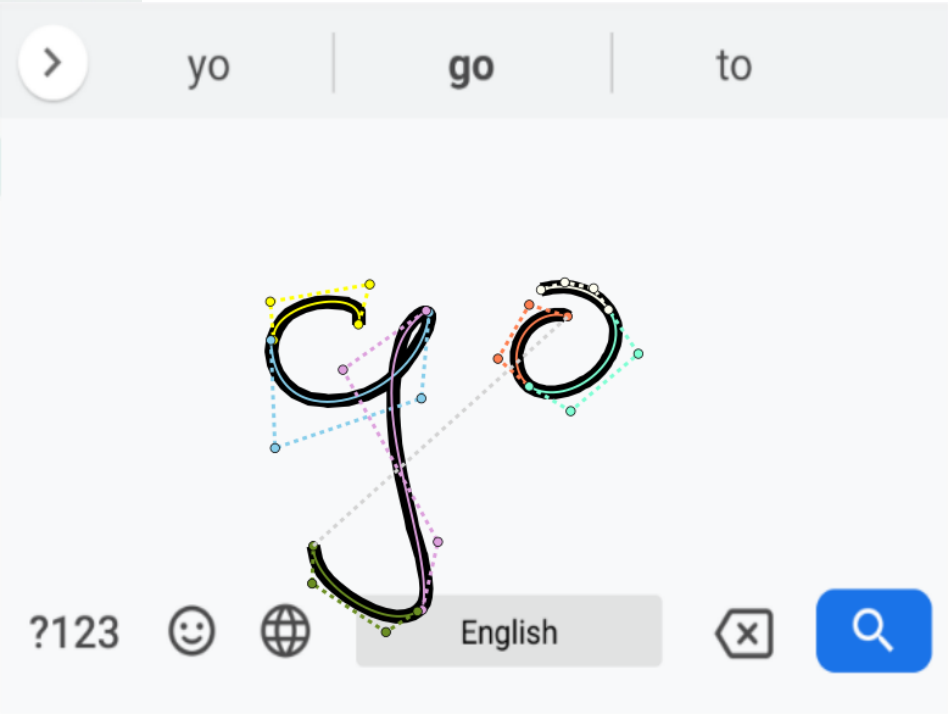
Guessing what you want to draw



GBoard

Guessing what you are writing





Game Over