

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

課號：Psy7277

識別碼：227M9280

教室：北 206

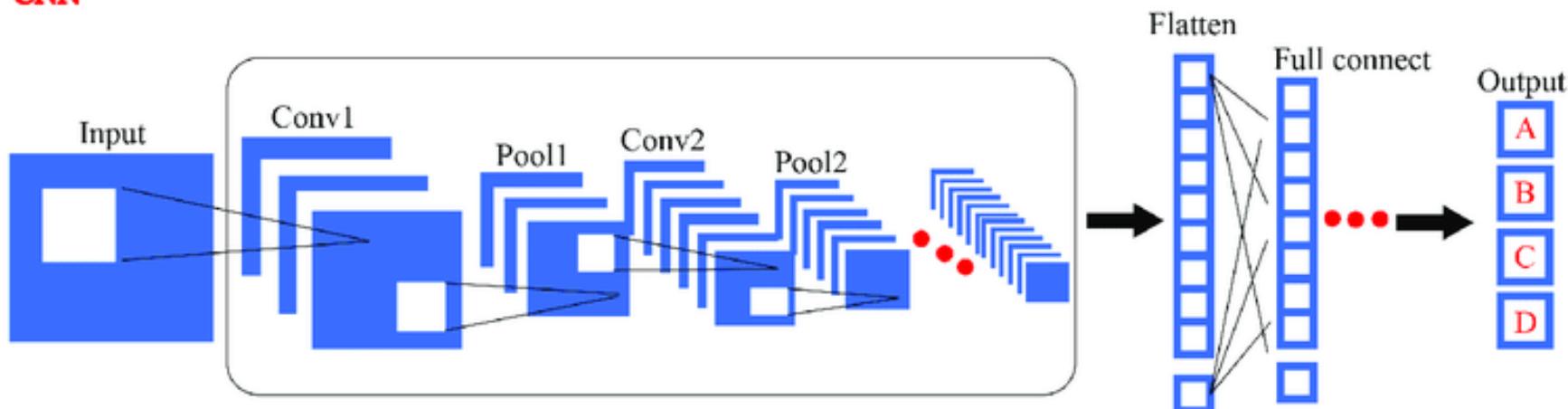
時間：五 234



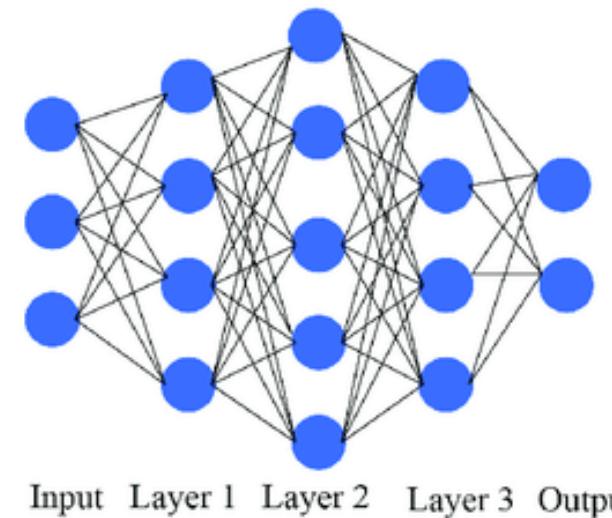
Deep Neural Networks 的分類

CNN 通常處理影像資料；RNN 通常處理語言資料

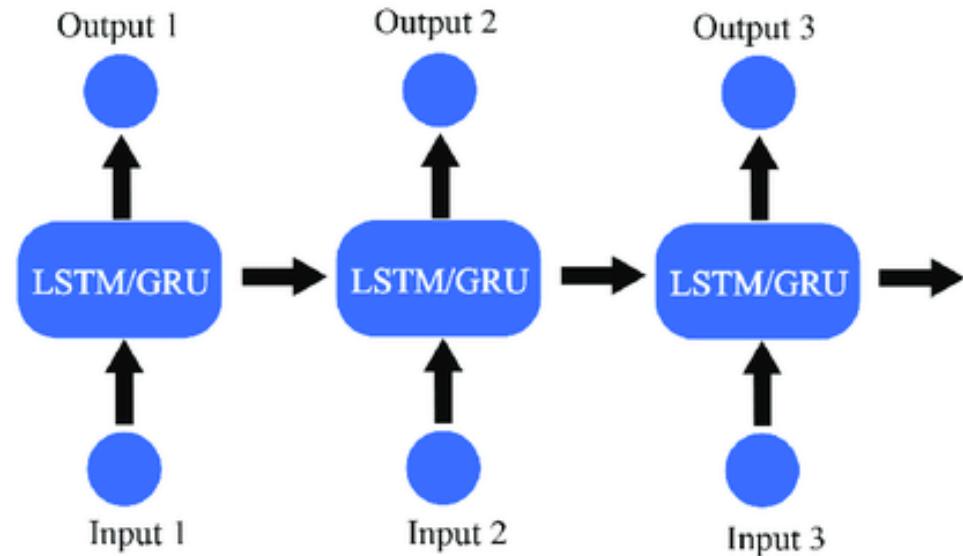
CNN



DNN

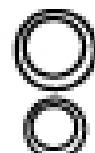


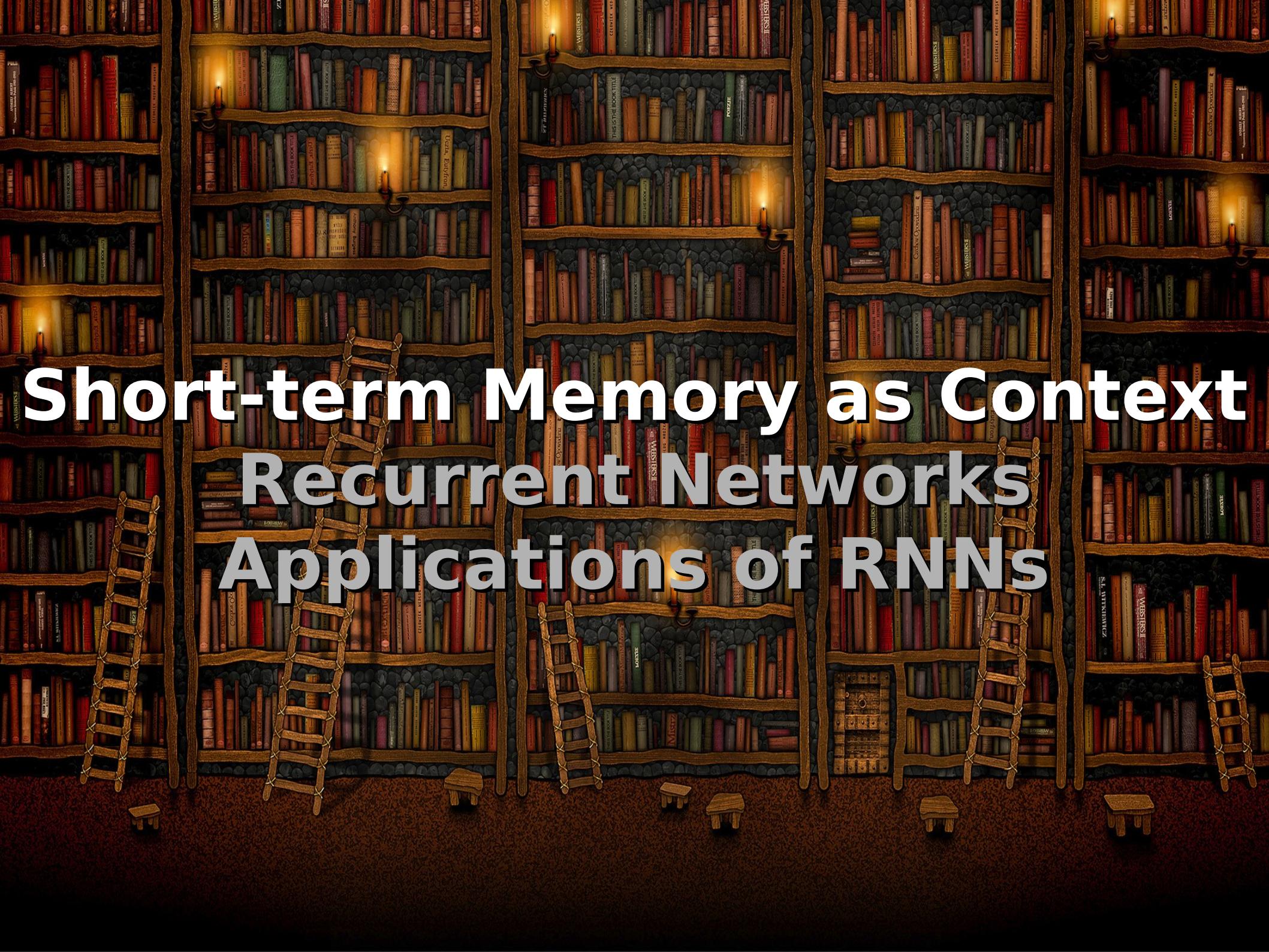
RNN





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

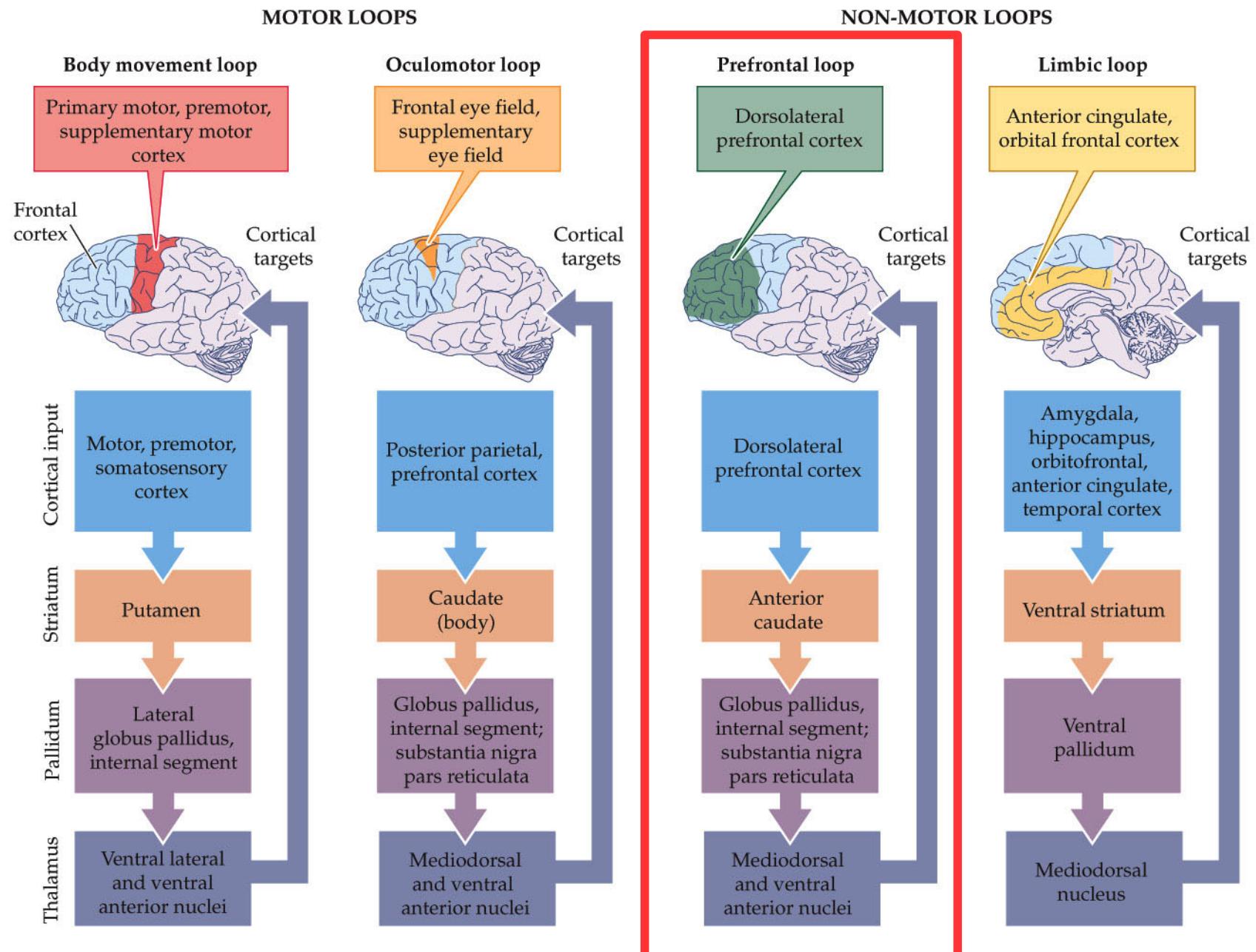
數學上無解：

$x = \text{紅} \rightarrow y = \text{紅}$ (叫意)
 $x = \text{紅} \rightarrow y = \text{綠}$ (叫色)

數學上有解：

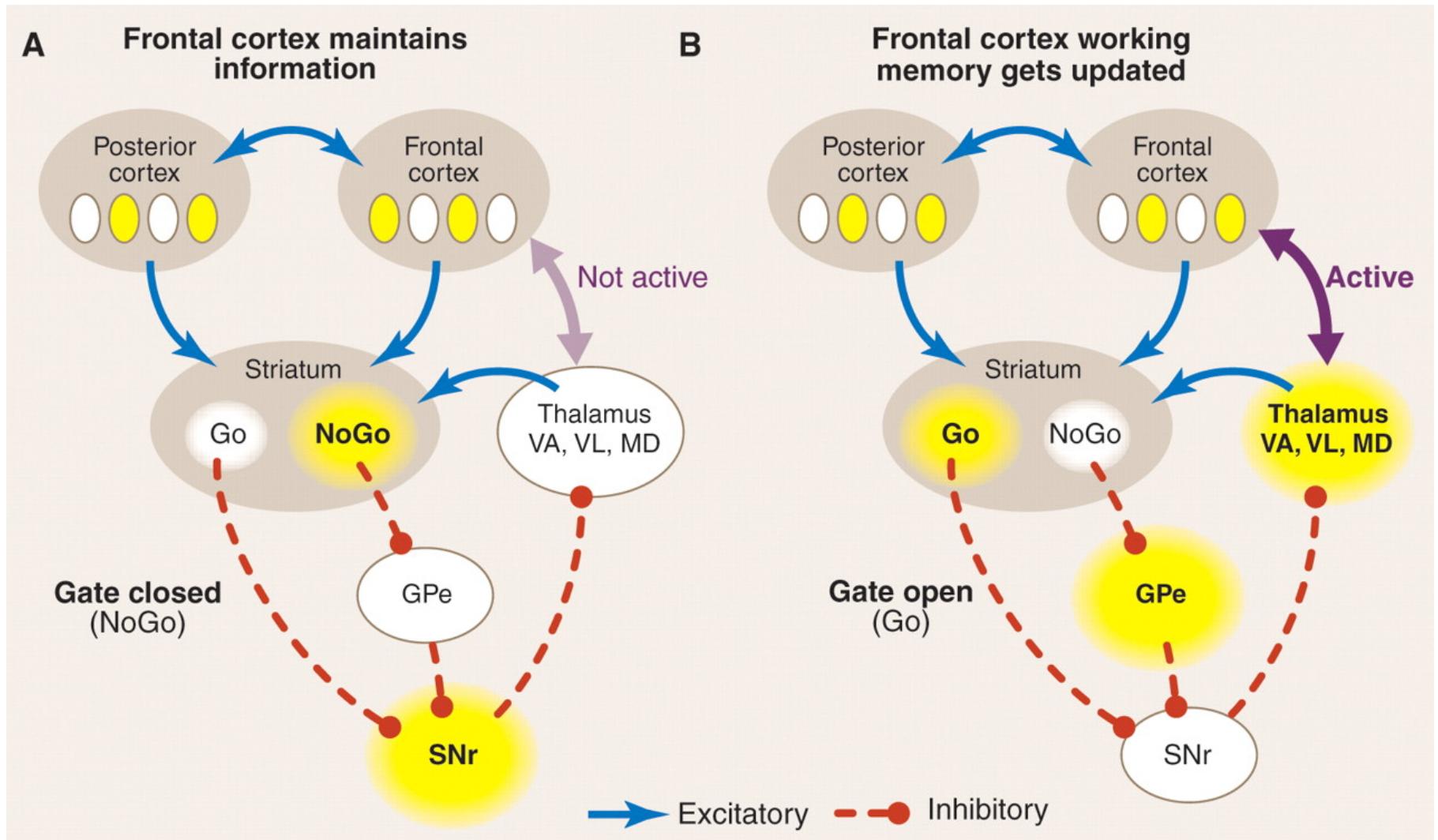
$x = (\text{紅}, \text{叫意}) \rightarrow y = \text{紅}$
 $x = (\text{紅}, \text{叫色}) \rightarrow y = \text{綠}$

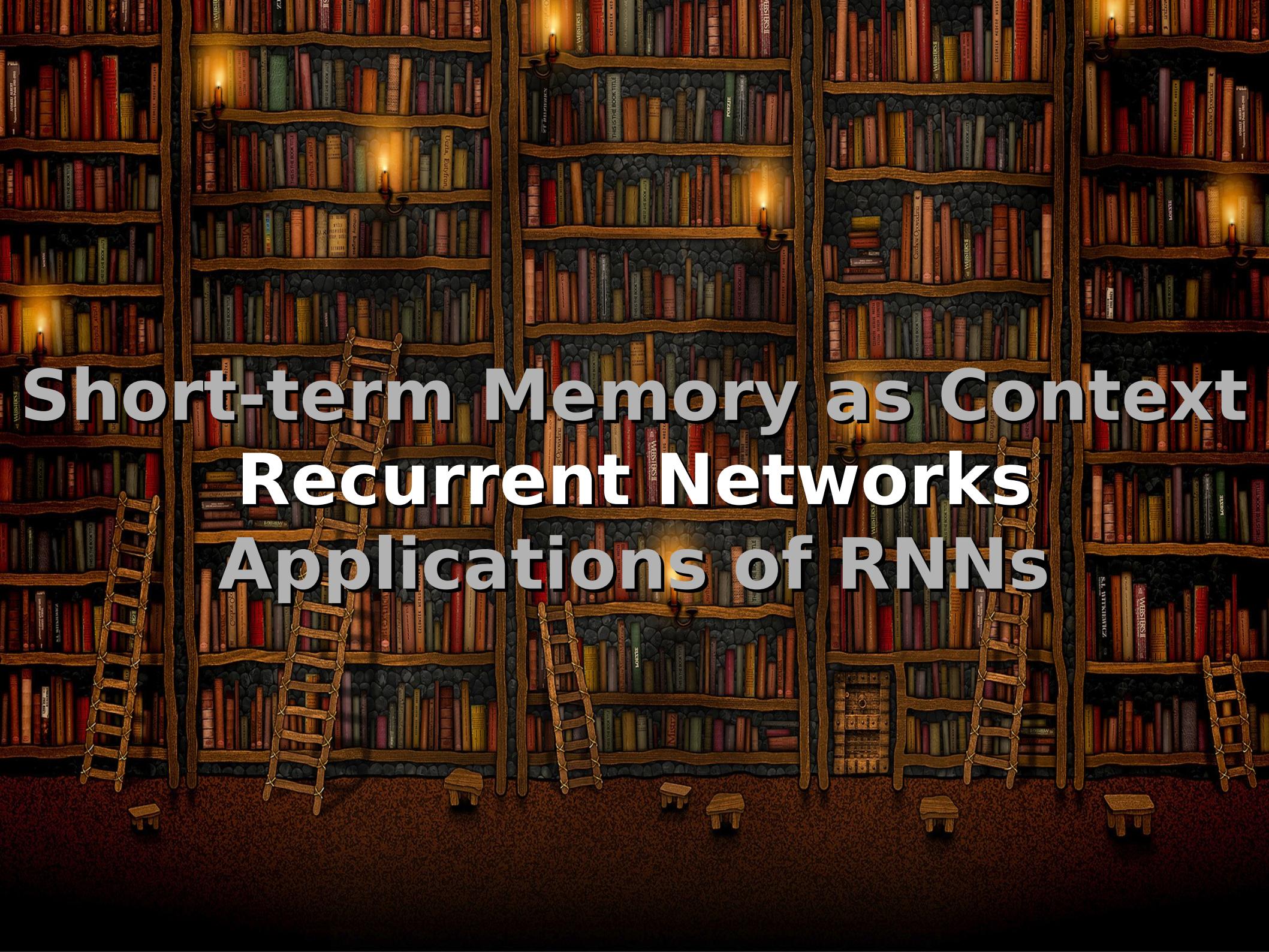
Cortico-Basal Ganglia Loops



BG-modulated PFC

= Prefrontal cortex Basal ganglia Working Memory

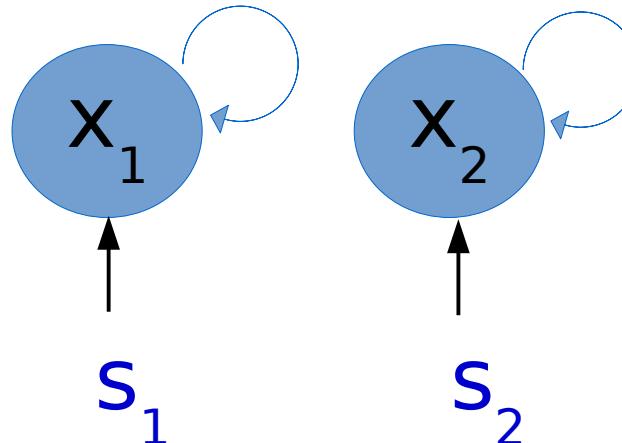




Short-term Memory as Context Recurrent Networks Applications of RNNs

Self-Recurrent Excitations

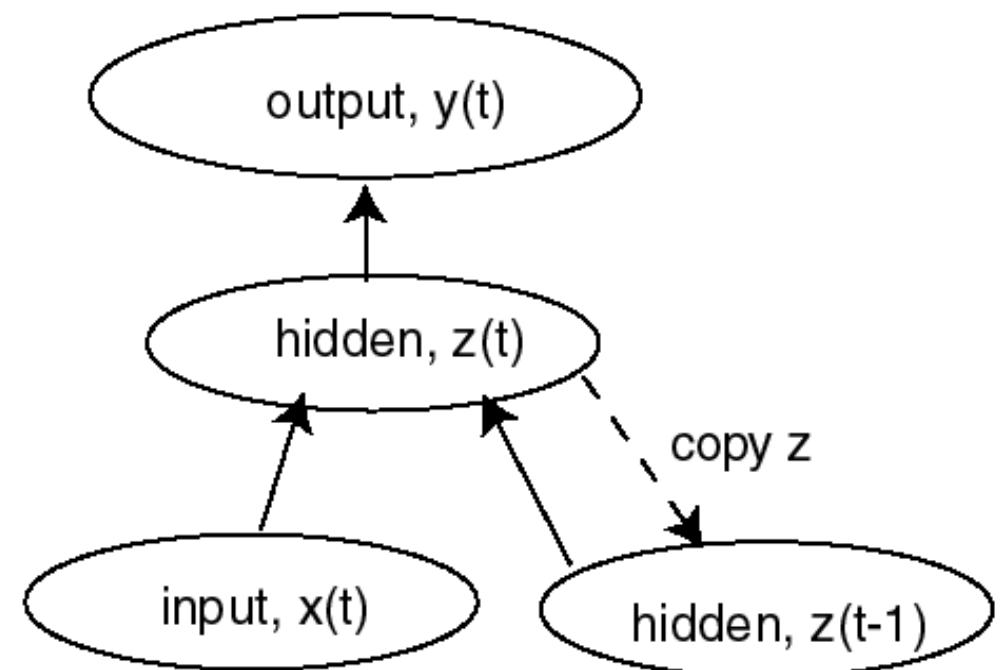
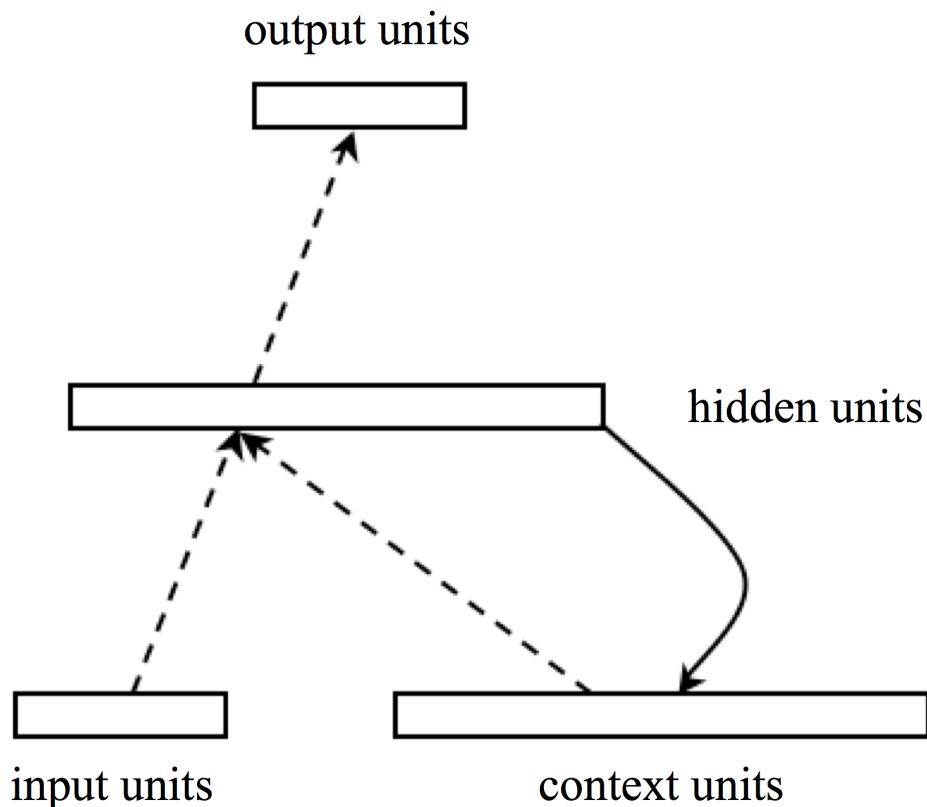
自我連結的刺激可維持反應卻無法維持刺激對比



```
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)

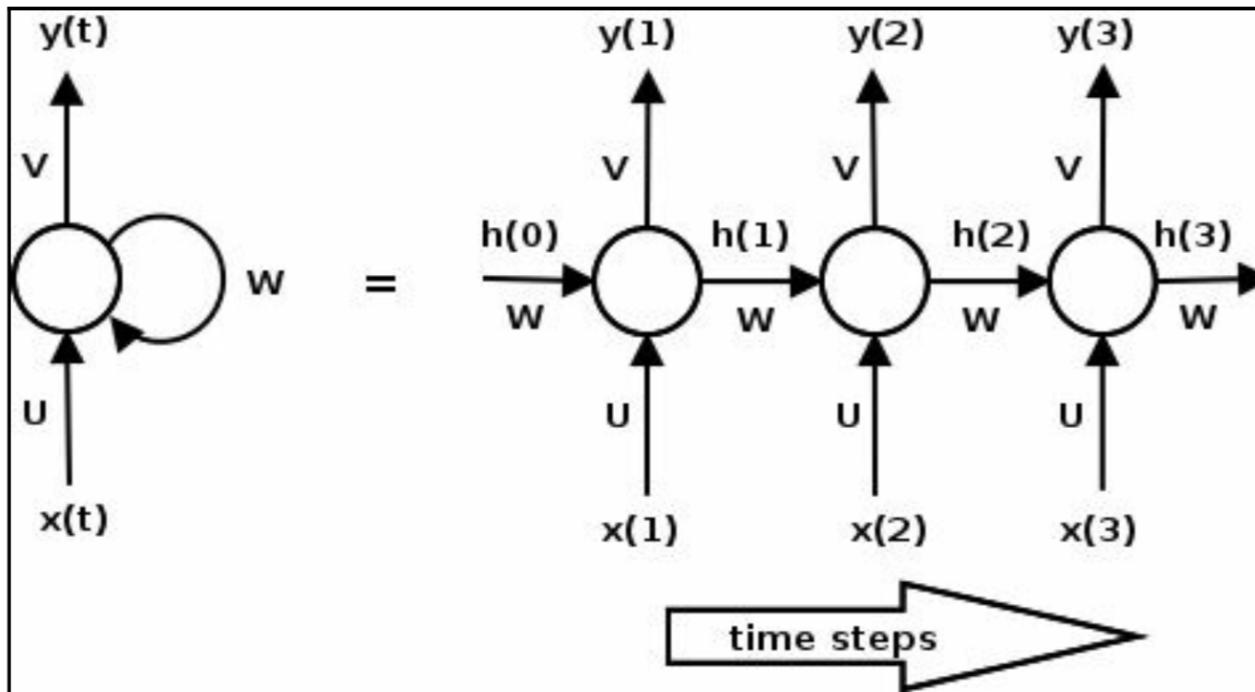
by Elman (1990)



to process sequential information

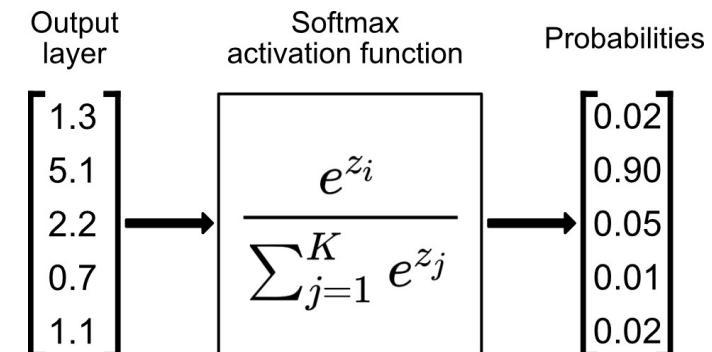
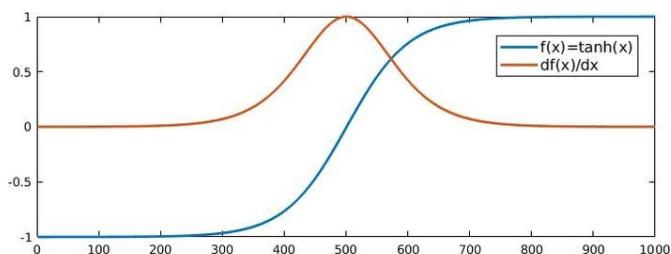
Simple RNN (2/2)

effectively has VERY deep layers



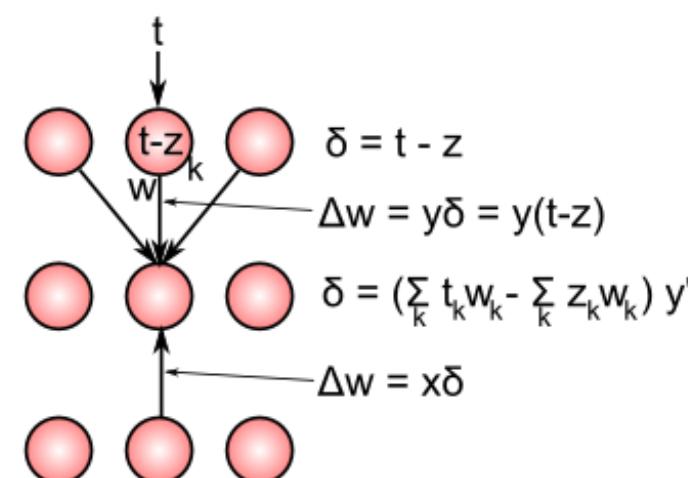
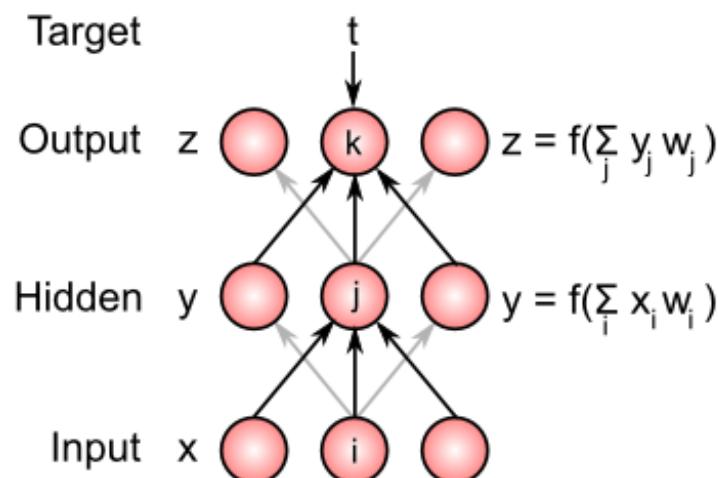
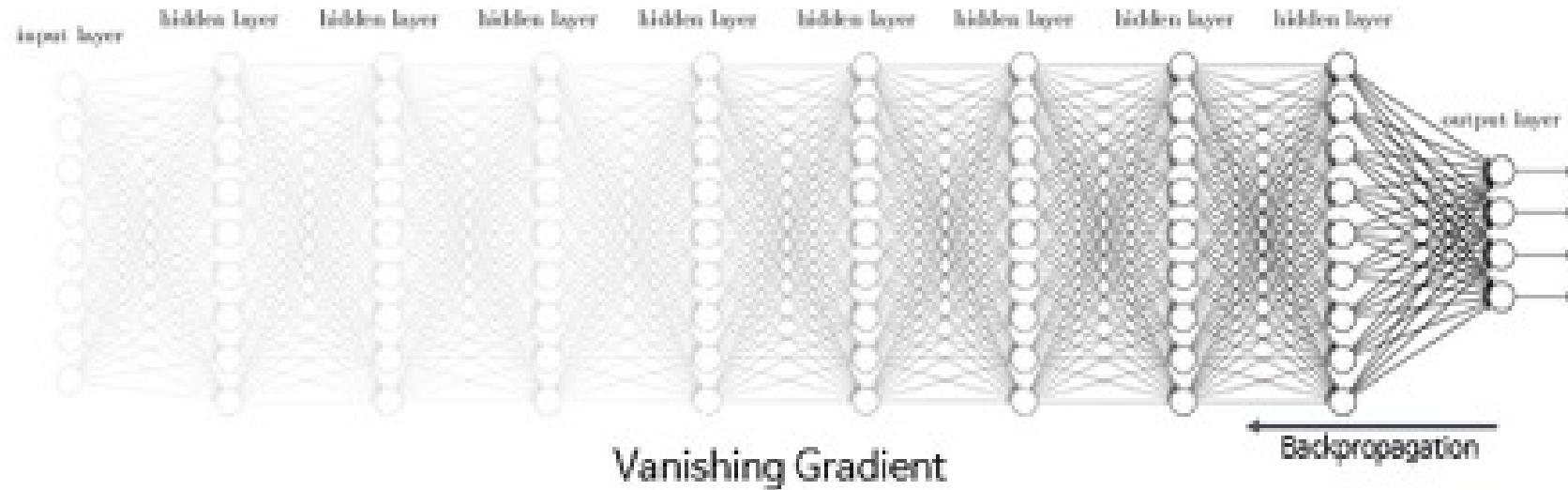
$$h_t = \tanh(W h_{t-1} + U x_t)$$

$$y_t = \text{softmax}(V h_t)$$

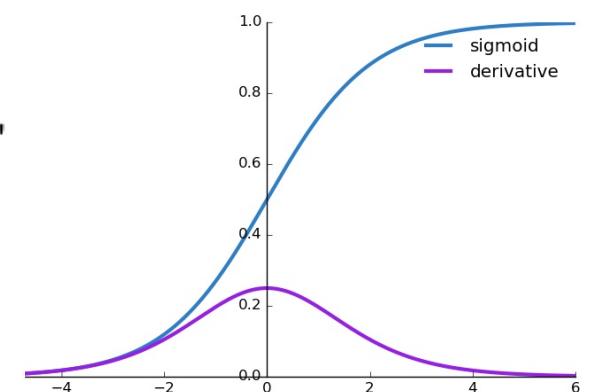


Temporal Backpropagation

suffers from vanishing & exploding gradient problem

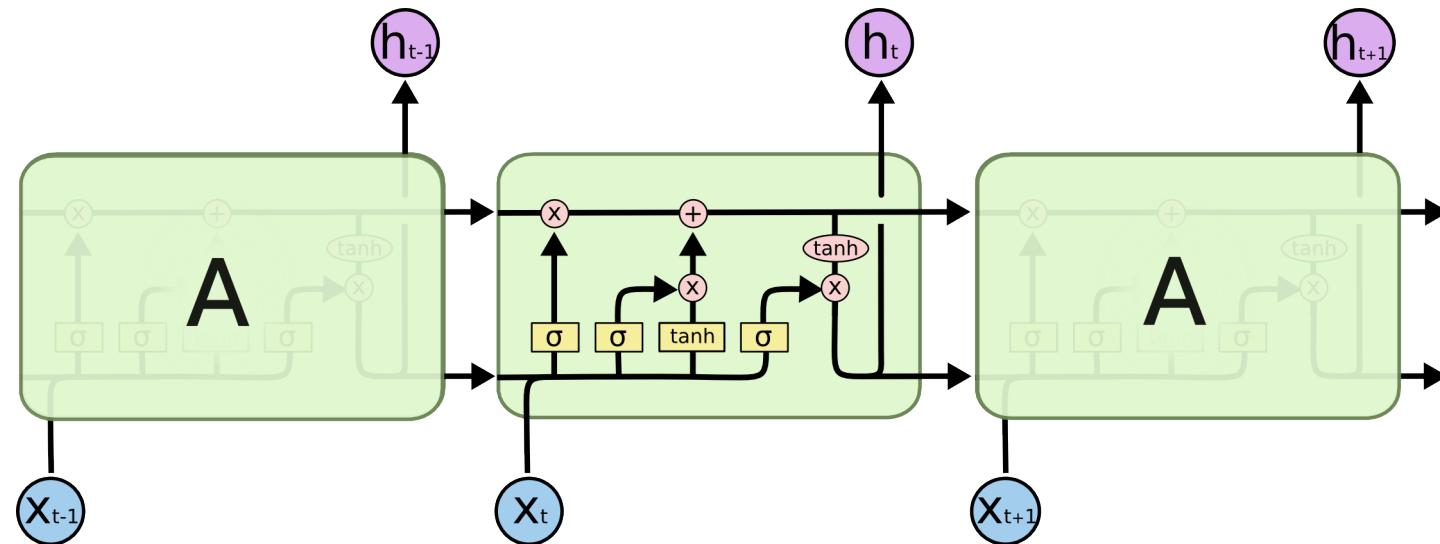
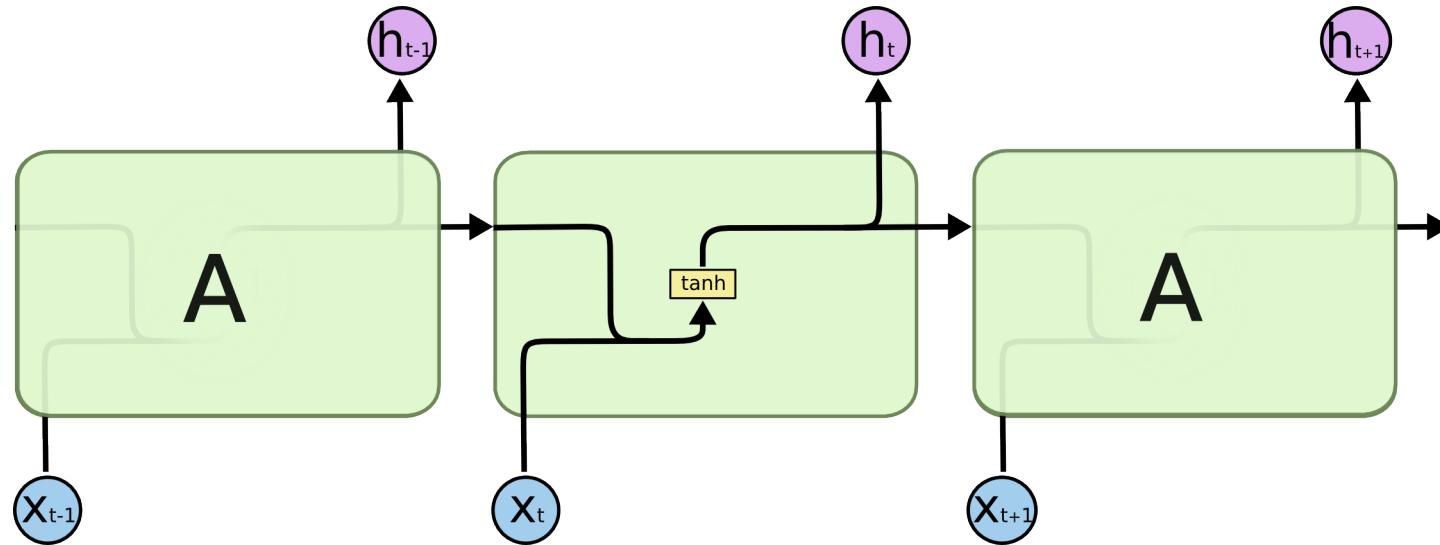


a) Feedforward Activation b) Error Backpropagation



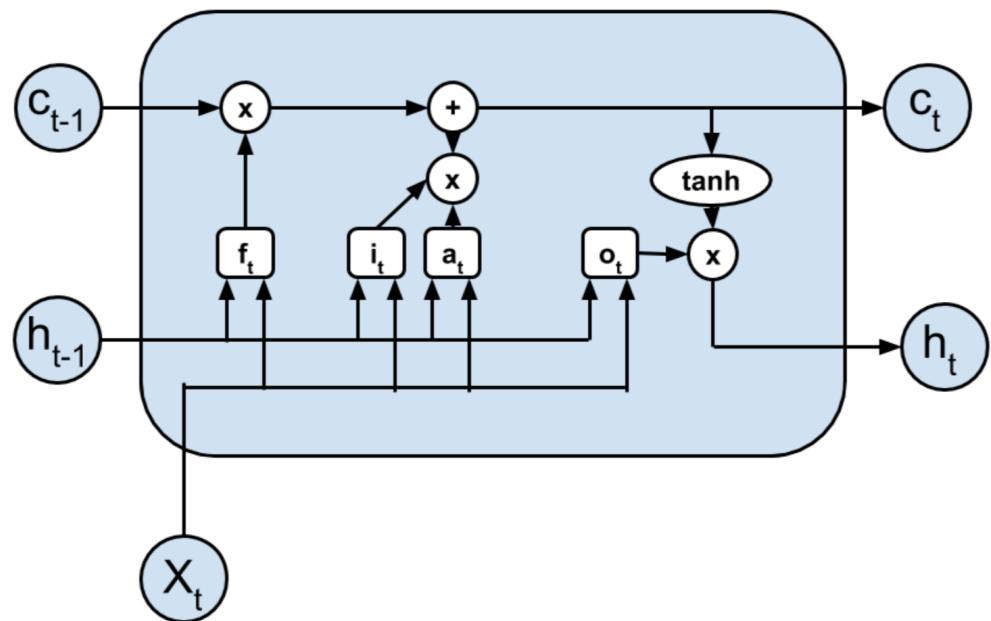
Long Short-term Memory (1/2)

introduces input, forget, output gates



Long Short-term Memory (2/2)

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



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

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

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

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

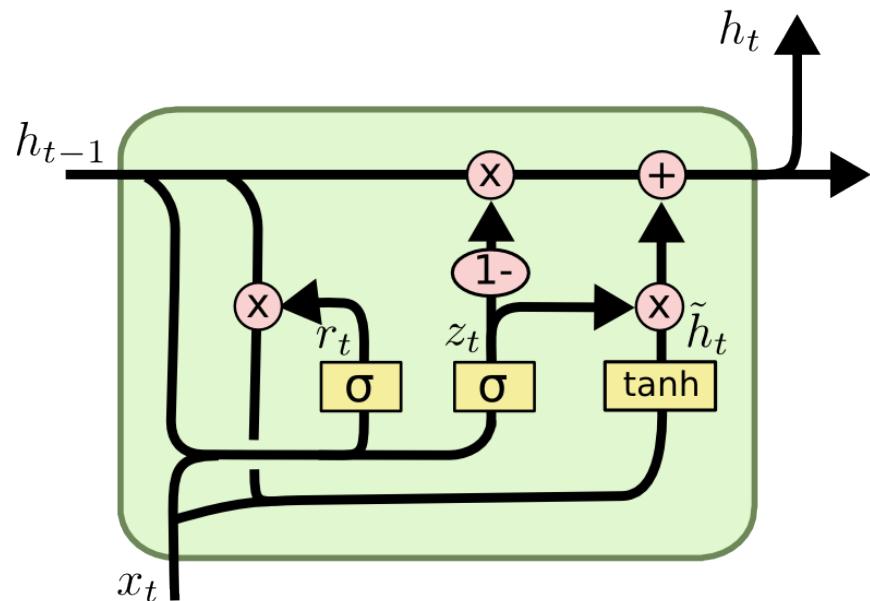
$$c_t = f_t * c_{t-1} + i_t * a_t$$

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

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

Gated Recurrent Unit

GRU is a simplified LSTM without an output gate



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

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

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

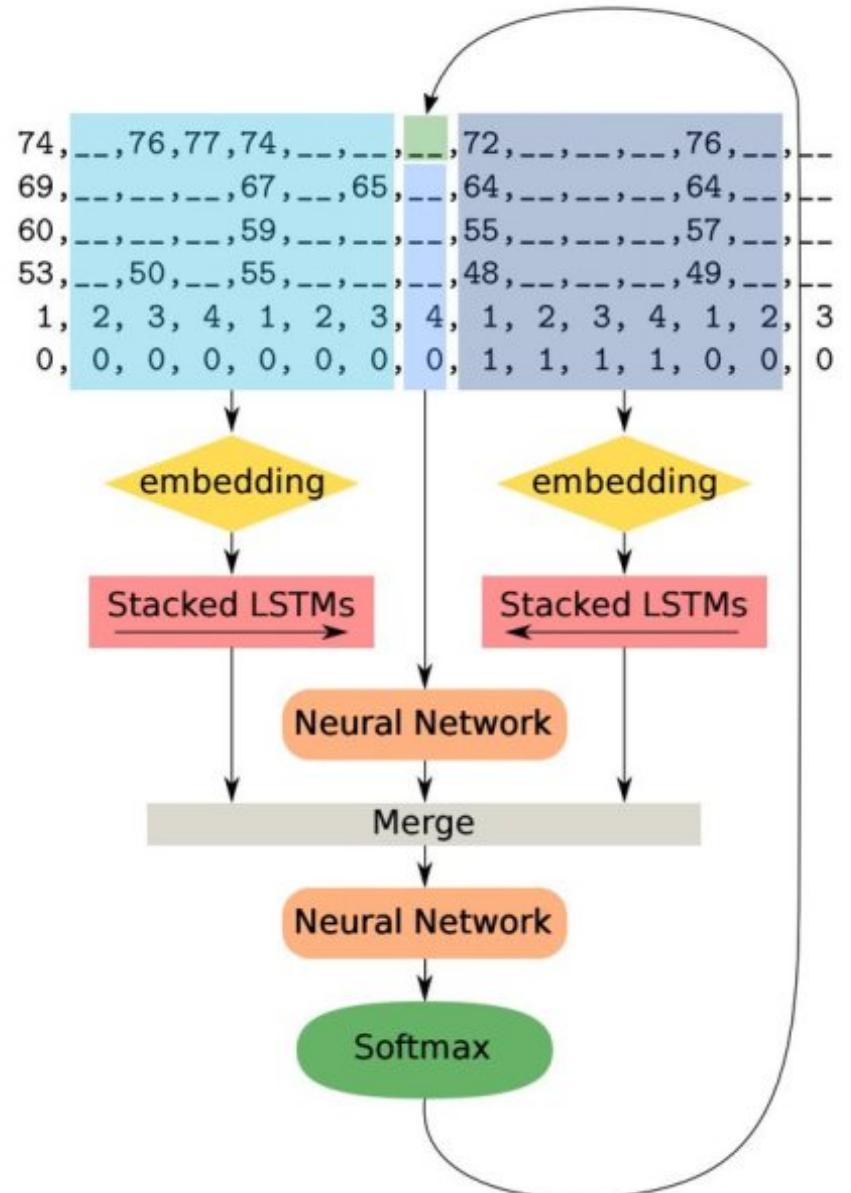
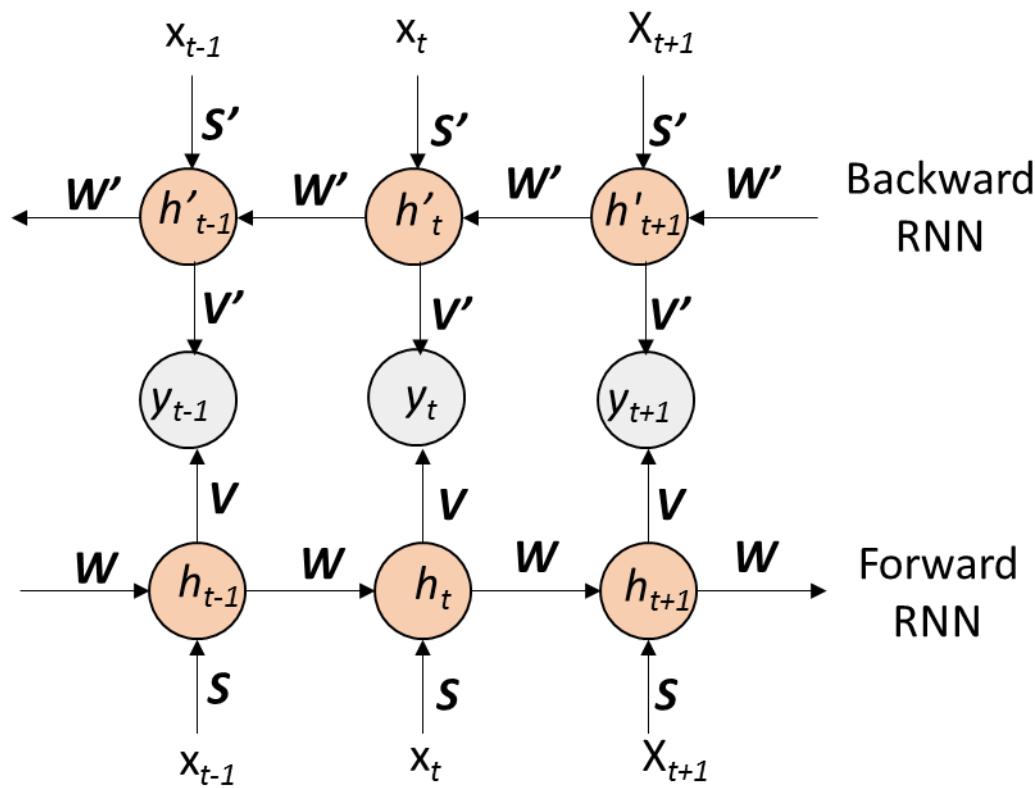
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

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

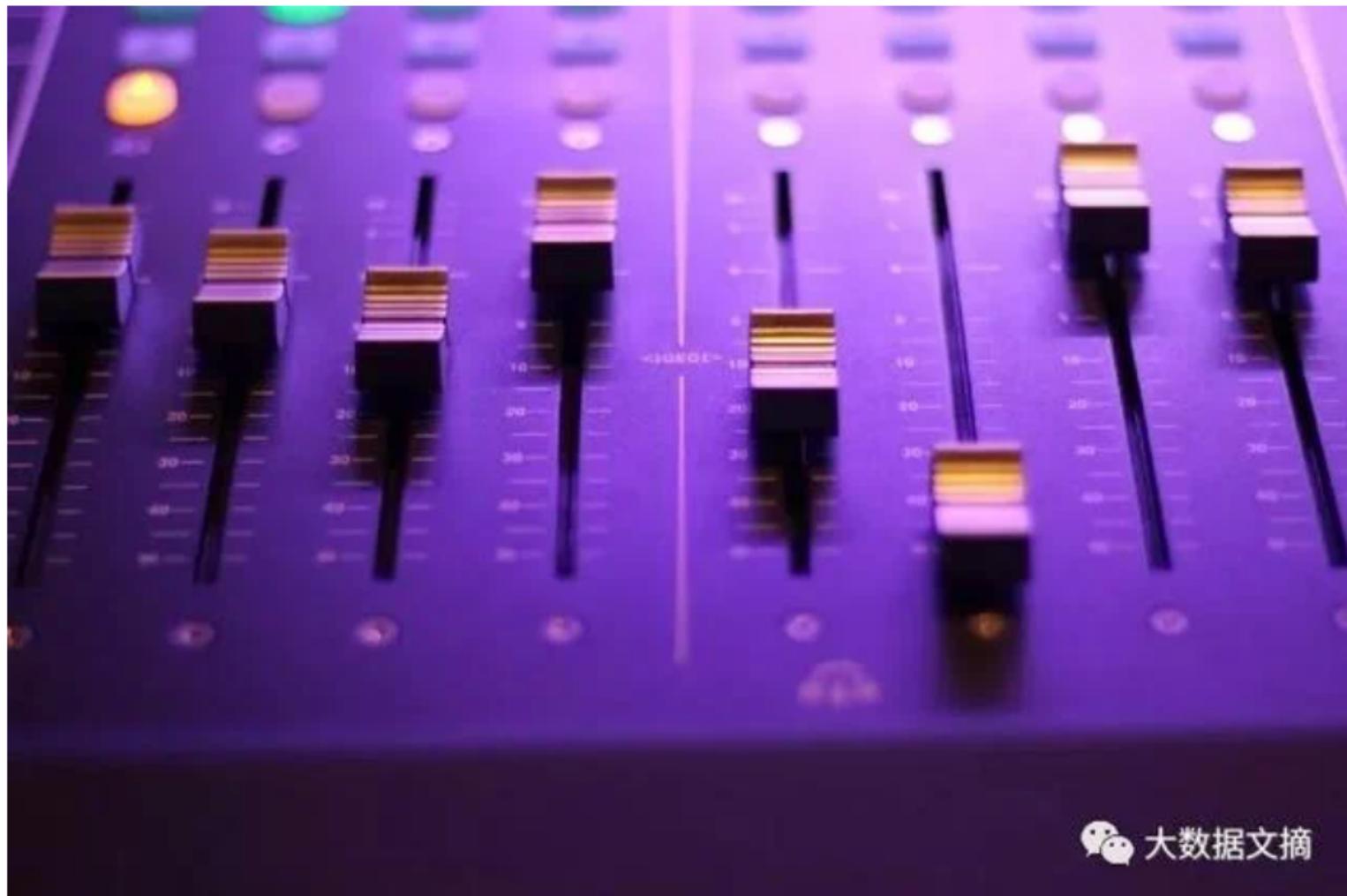
Abstract

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)

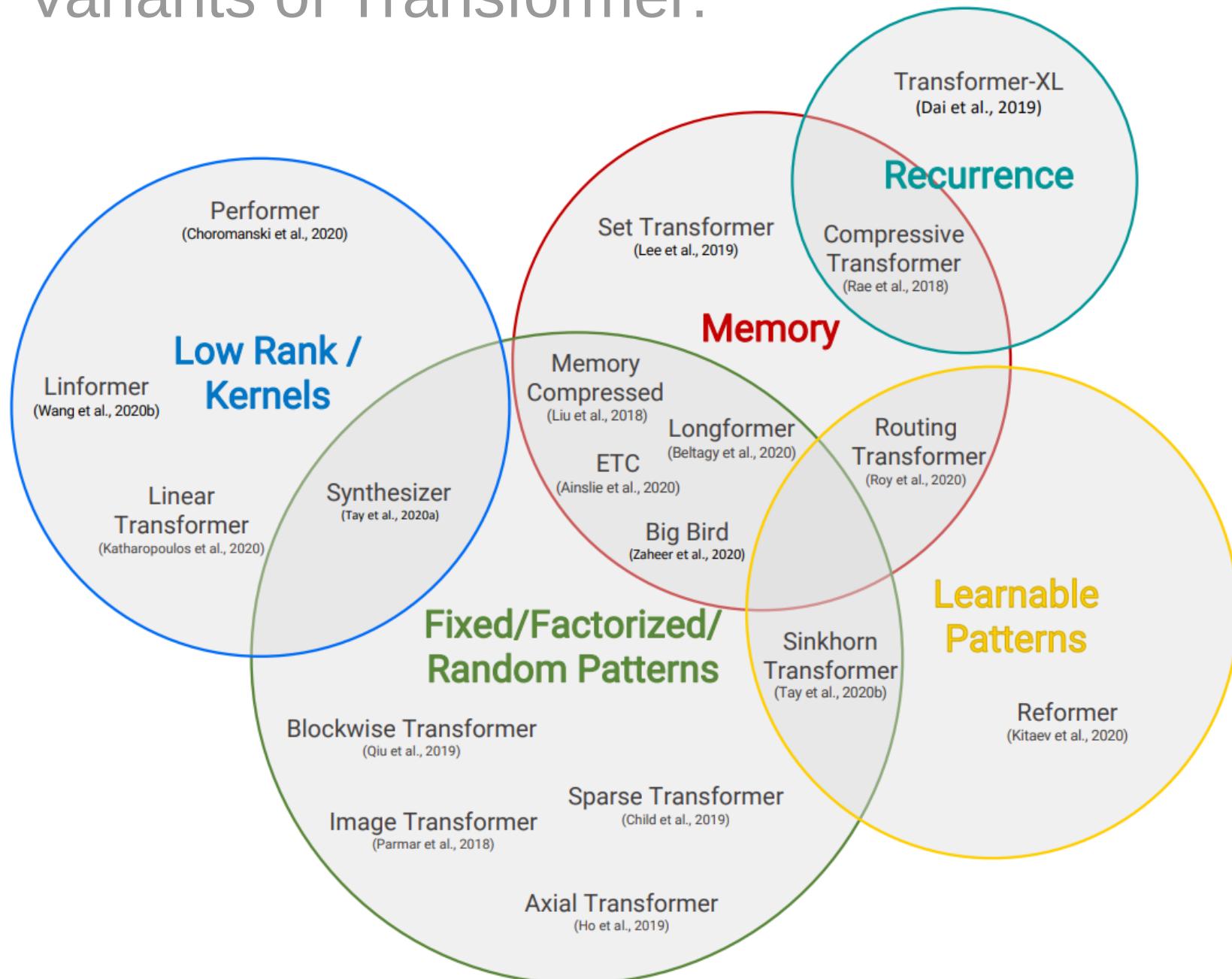
別再用RNN和LSTM了！注意力模型才是王道

2018-05-02 由 大數據文摘 發表于科技

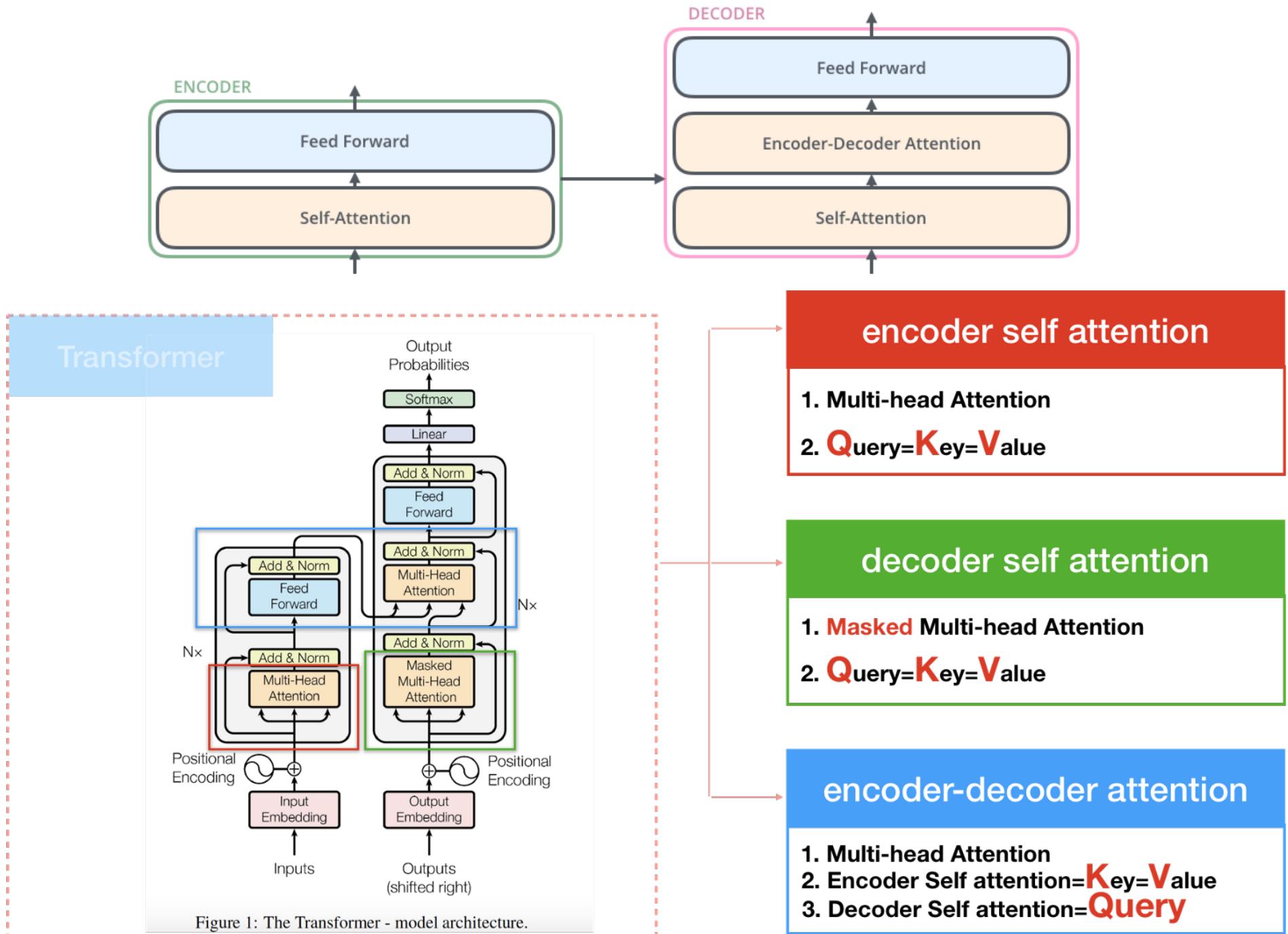


Attention replacing RNN (3/3)

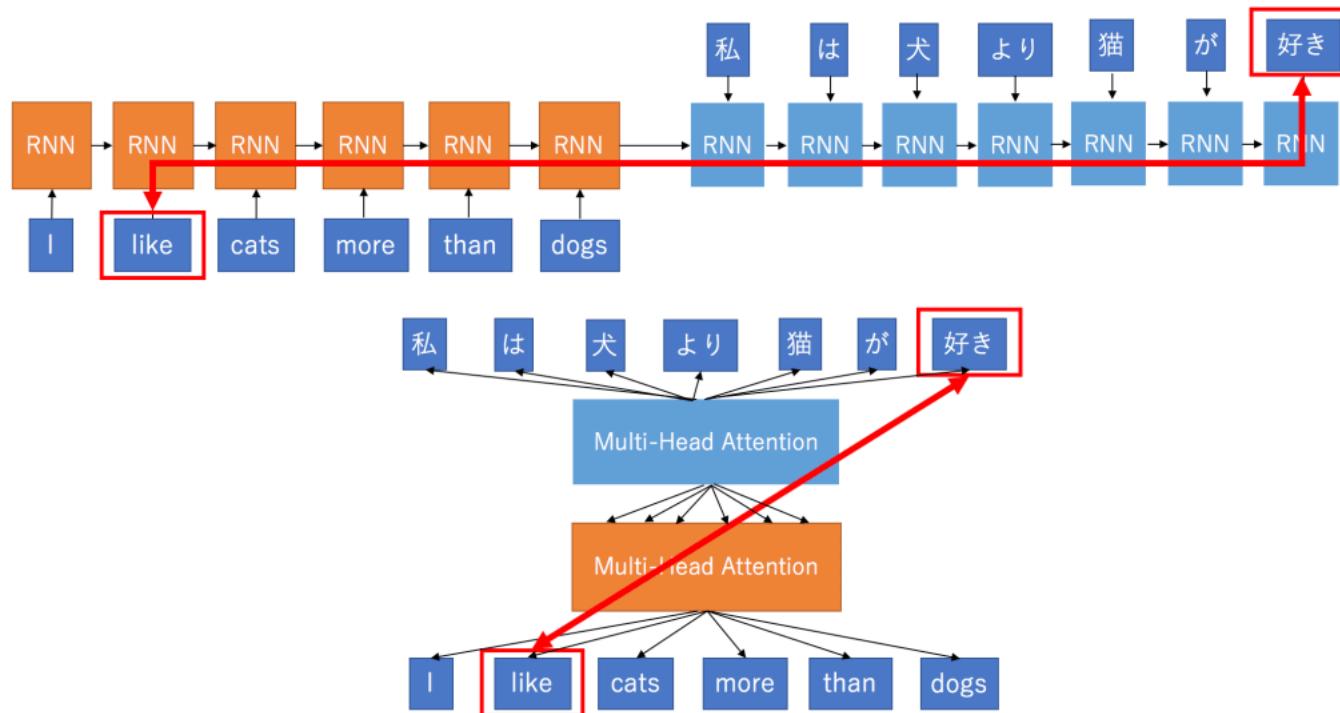
Variants of Transformer:



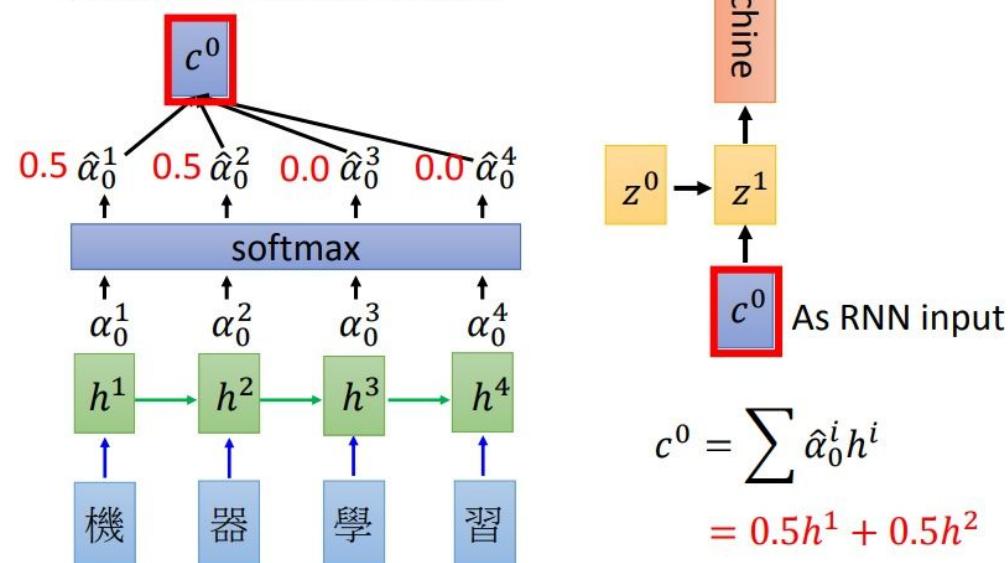
Attention in Transformer



Encoder-Decoder Attention



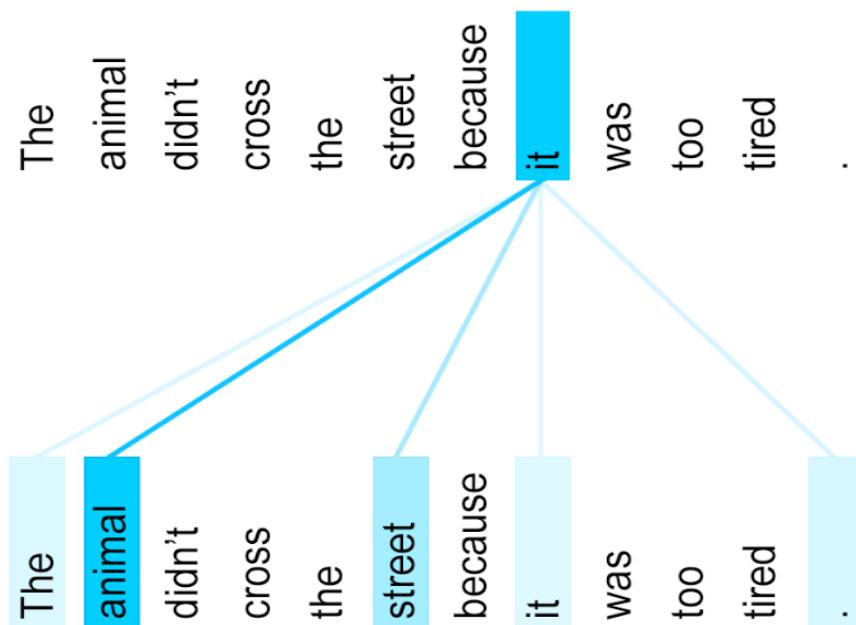
- Attention-based model



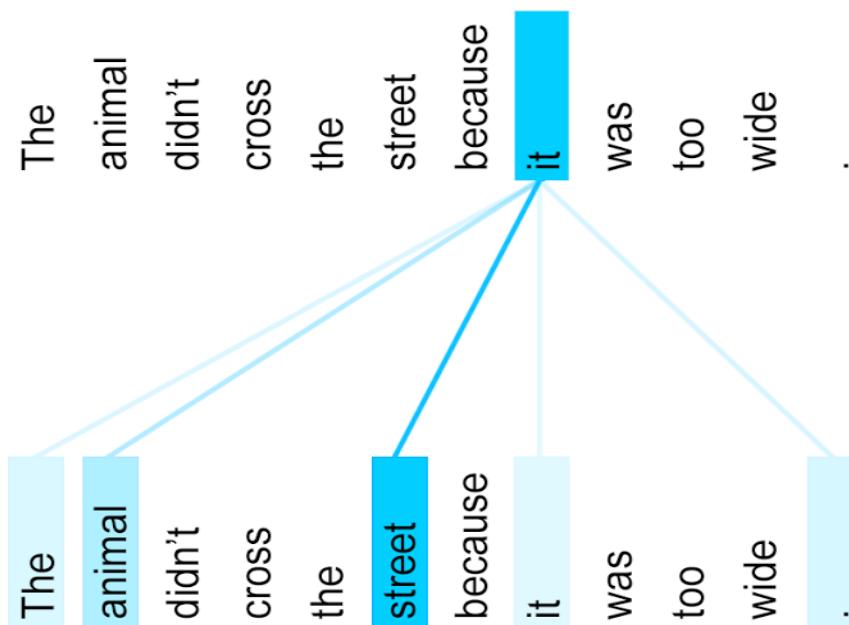
Encode/Decoder Self-Attention

Example: What's the correct understanding of “it”?

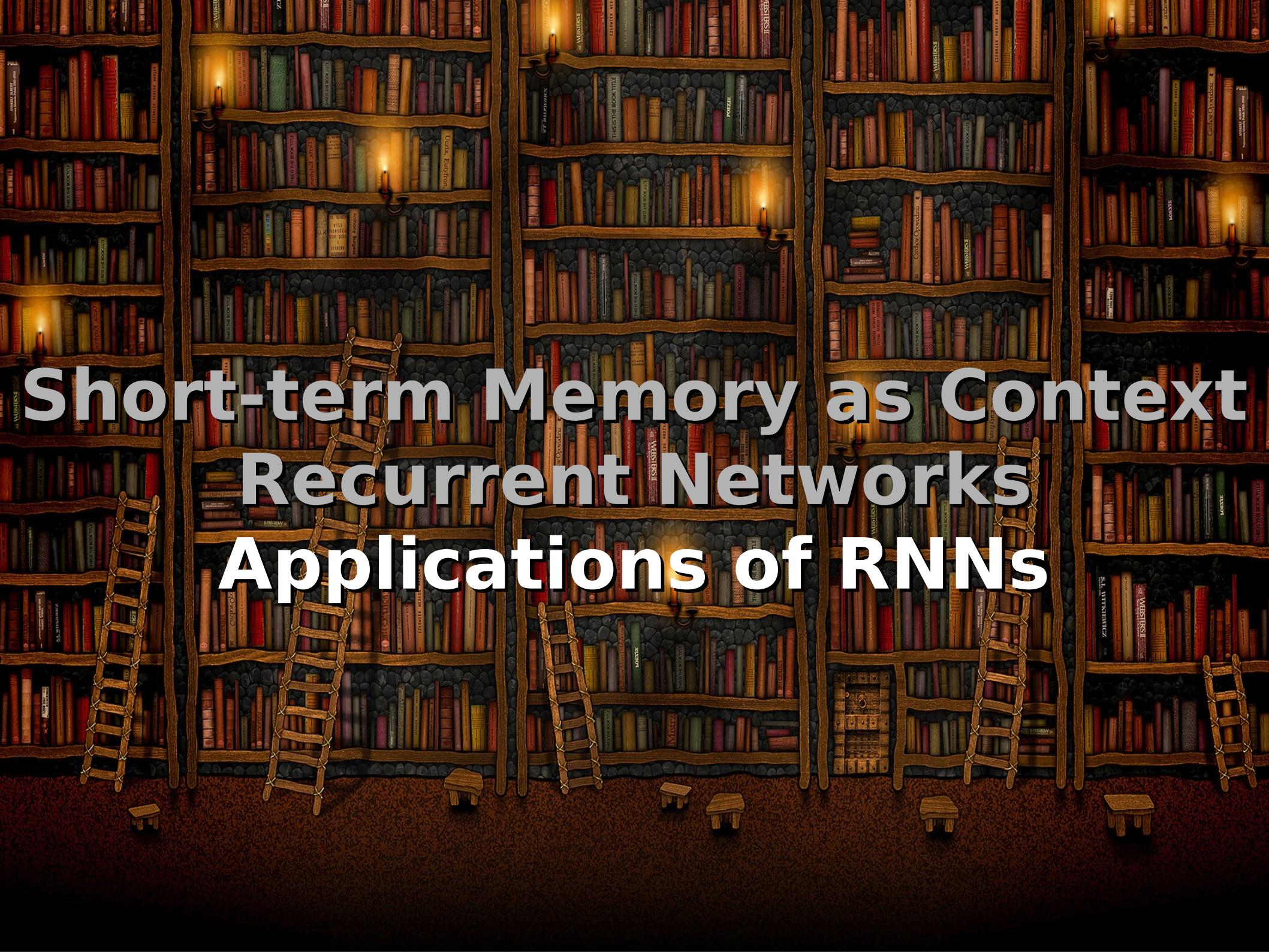
The animal didn't cross the street because it was too tired .



The animal didn't cross the street because it was too wide .



Different understanding of “it” = Different vector of “it”

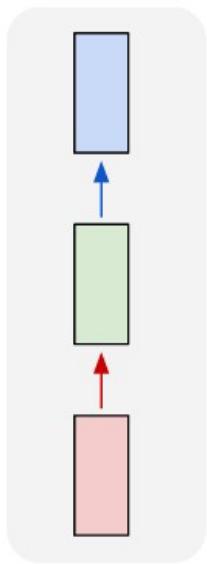


Short-term Memory as Context Recurrent Networks Applications of RNNs

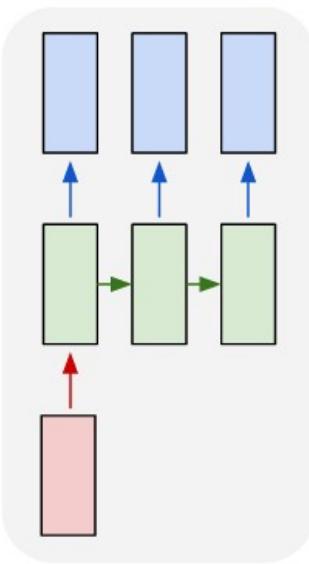
Recurrent Neural Networks

take many forms

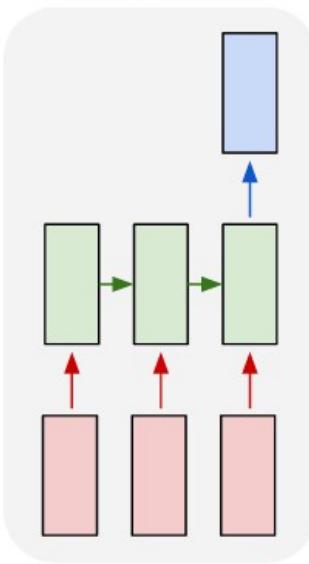
one to one



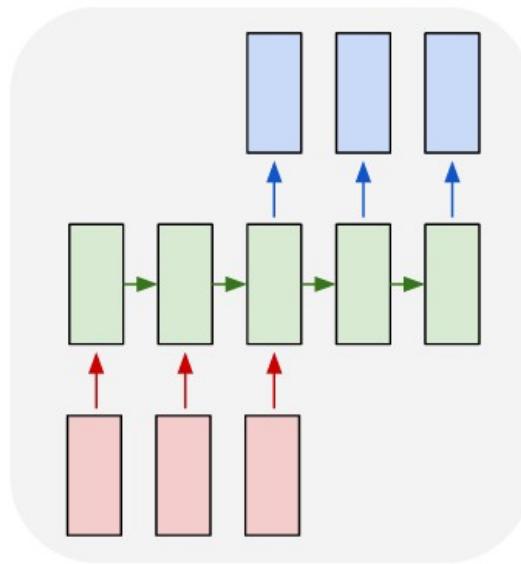
one to many



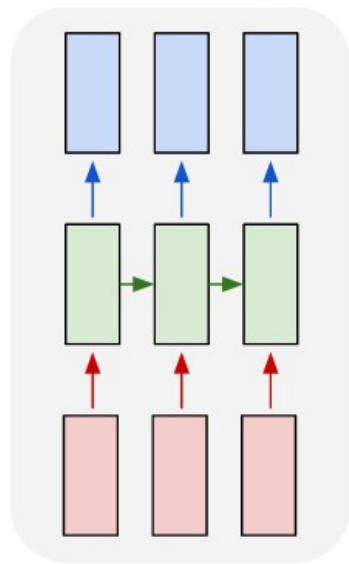
many to one



many to many



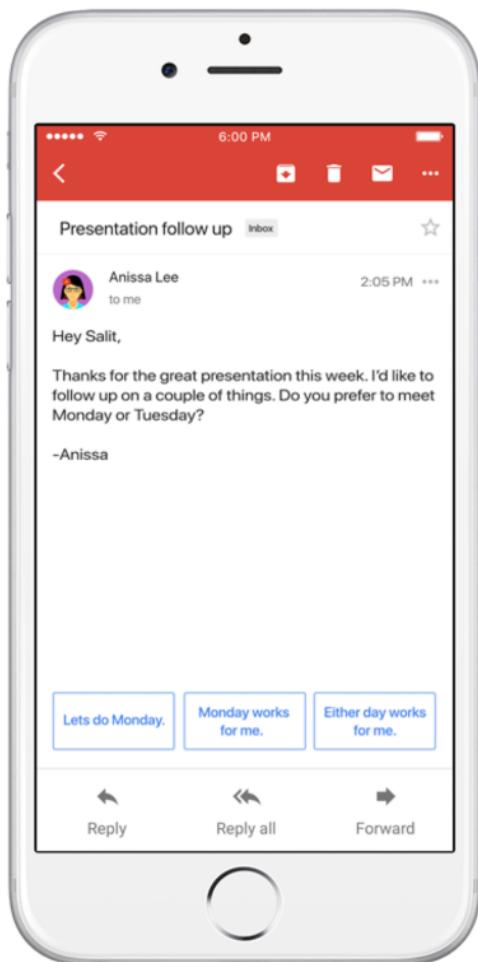
many to many



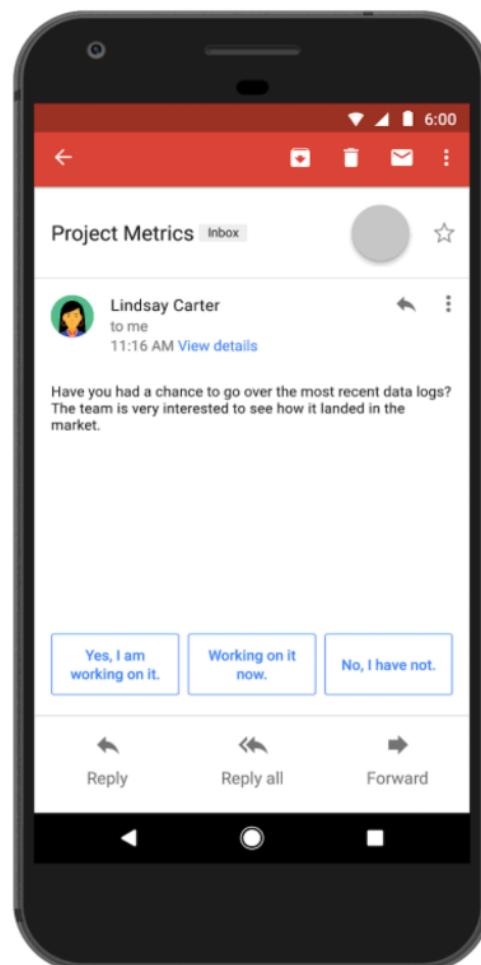
for image classification, image captioning,
sentiment classification, language translation,
speech recognition, etc.

Google SmartReply

Guessing what you want to reply

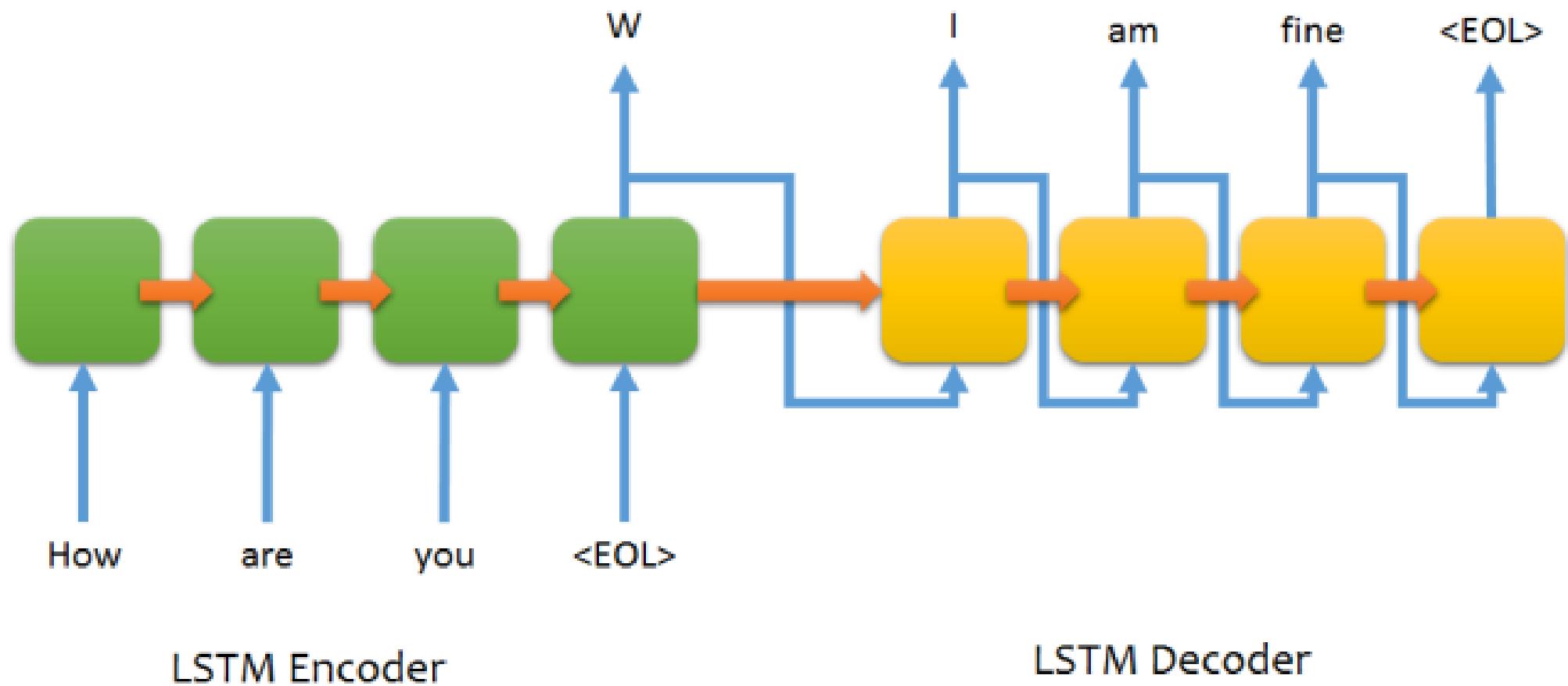


Smart Reply in Gmail



Chatbots (using Seq2Seq)

Guessing what you want to know

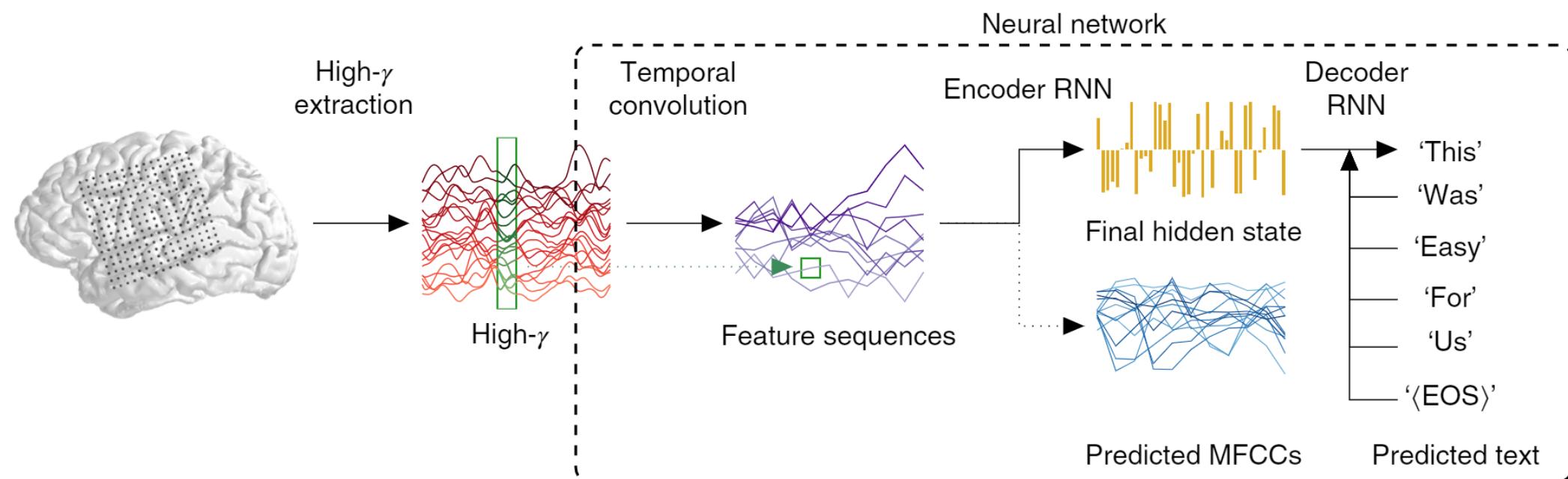


Data Modeling 範例

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

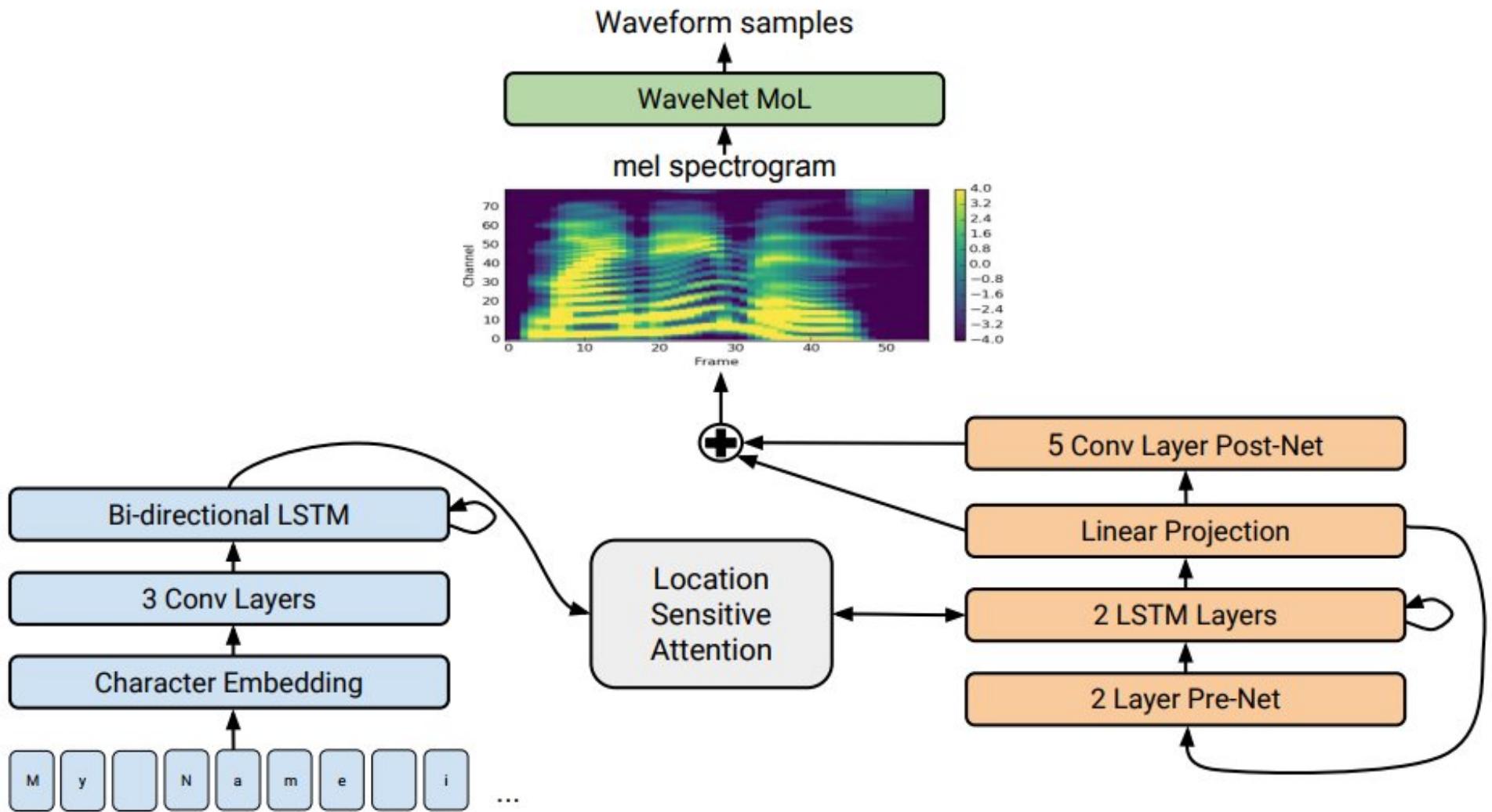
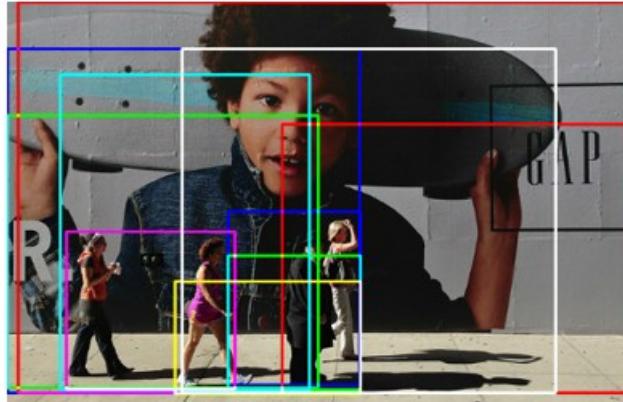


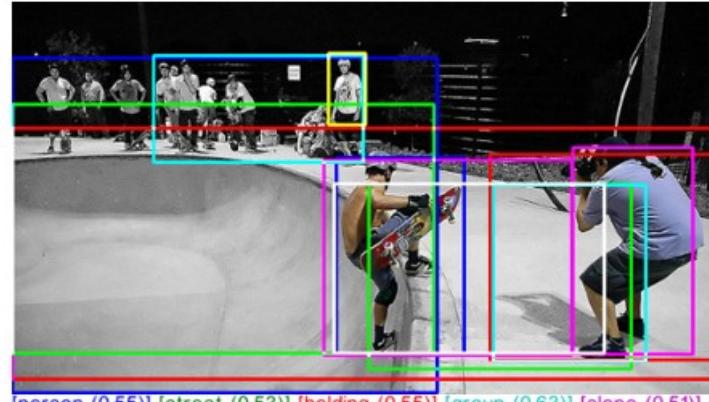
Image Captioning

Image to text



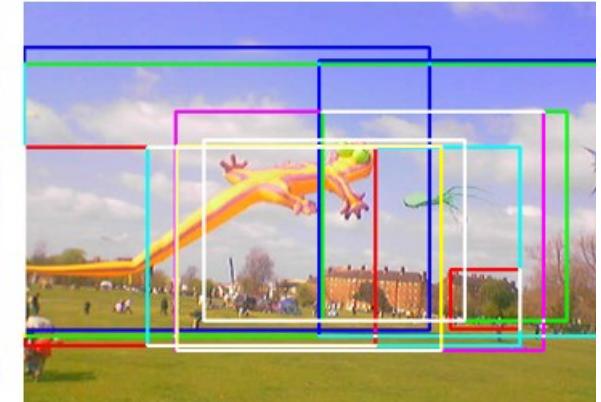
[men (0.59)] [group (0.66)] [woman (0.64)]
[people (0.89)] [holding (0.60)] [playing (0.61)] [tennis (0.69)]
[court (0.51)] [standing (0.59)] [skis (0.58)] [street (0.52)]
[man (0.77)] [skateboard (0.67)]

a group of people standing next to each other
people stand outside a large ad for gap featuring a young boy



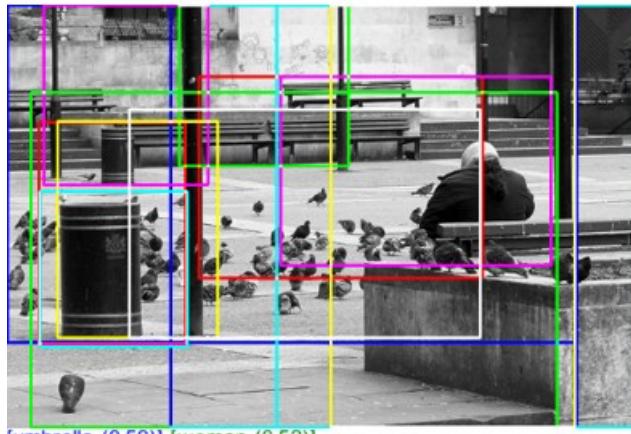
[person (0.55)] [street (0.53)] [holding (0.55)] [group (0.63)] [slope (0.51)]
[standing (0.62)] [snow (0.91)] [skis (0.74)] [player (0.54)]
[people (0.85)] [men (0.57)] [skiing (0.51)]
[skateboard (0.89)] [riding (0.75)] [tennis (0.74)] [trick (0.53)] [skate (0.52)]
[woman (0.52)] [man (0.86)] [down (0.61)]

a group of people riding skis down a snow covered slope
a guy on a skate board on the side of a ramp



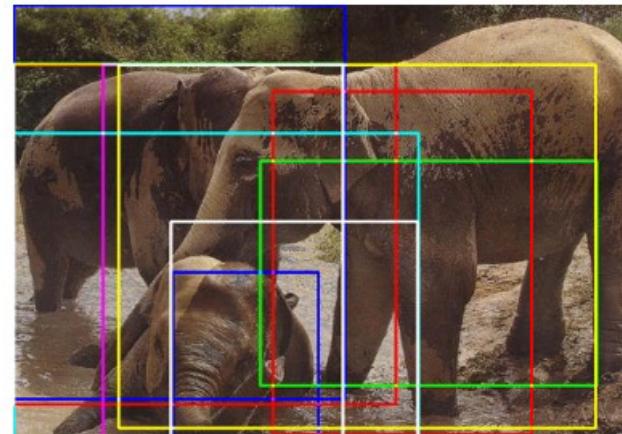
[airplane (0.57)] [plane (0.58)] [kites (0.93)] [people (0.80)]
[flying (0.93)] [man (0.57)] [beach (0.84)] [wave (0.61)]
[sky (0.61)] [kite (0.74)] [field (0.75)]

a couple of people flying kites in a field
people in a field flying different styles of kite



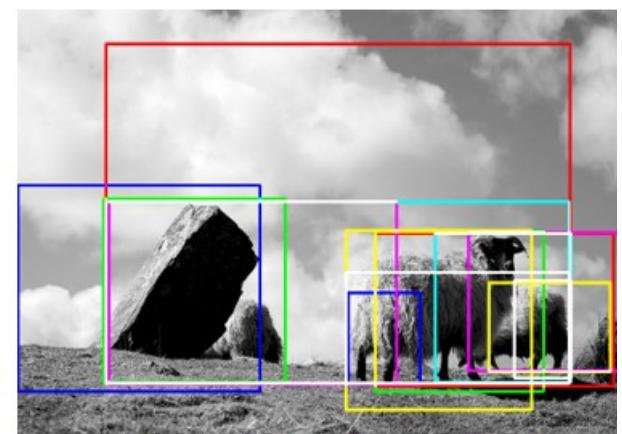
[umbrella (0.59)] [woman (0.52)]
[fire (0.96)] [hydrant (0.96)] [street (0.79)] [old (0.50)]
[bench (0.81)] [building (0.75)] [standing (0.57)] [baseball (0.55)]
[white (0.82)] [sitting (0.65)] [people (0.79)] [photo (0.53)]
[black (0.84)] [kitchen (0.54)] [man (0.72)] [water (0.56)]

a black and white photo of a fire hydrant
a courtyard full of poles pigeons and garbage cans also has benches on



[horse (0.53)] [bear (0.71)] [elephant (0.99)] [elephants (0.95)]
[brown (0.68)] [baby (0.62)] [walking (0.57)] [laying (0.61)]
[man (0.57)] [standing (0.79)] [field (0.65)]
[water (0.83)] [large (0.71)] [dirt (0.65)] [river (0.58)]

a baby elephant standing next to each other on a field



[man (0.59)] [beach (0.54)] [sky (0.53)] [bird (0.50)] [field (0.54)]
[snow (0.86)] [mountain (0.59)] [standing (0.81)] [white (0.64)]
[people (0.51)] [dog (0.60)] [cows (0.55)]
[sheep (0.97)] [black (0.84)] [grass (0.64)] [horse (0.60)]
[elephants (0.57)] [bear (0.81)]

a black bear standing on top of a grass covered

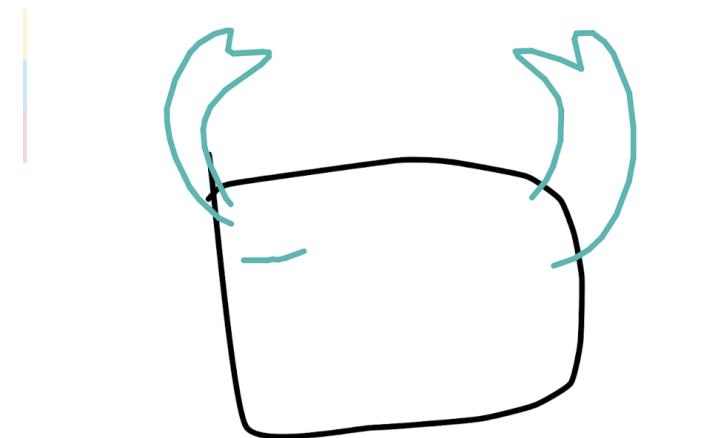
Sketch-RNN

Guessing what you want to draw

 info  random  clear 

Model: ▾

start drawing crab.

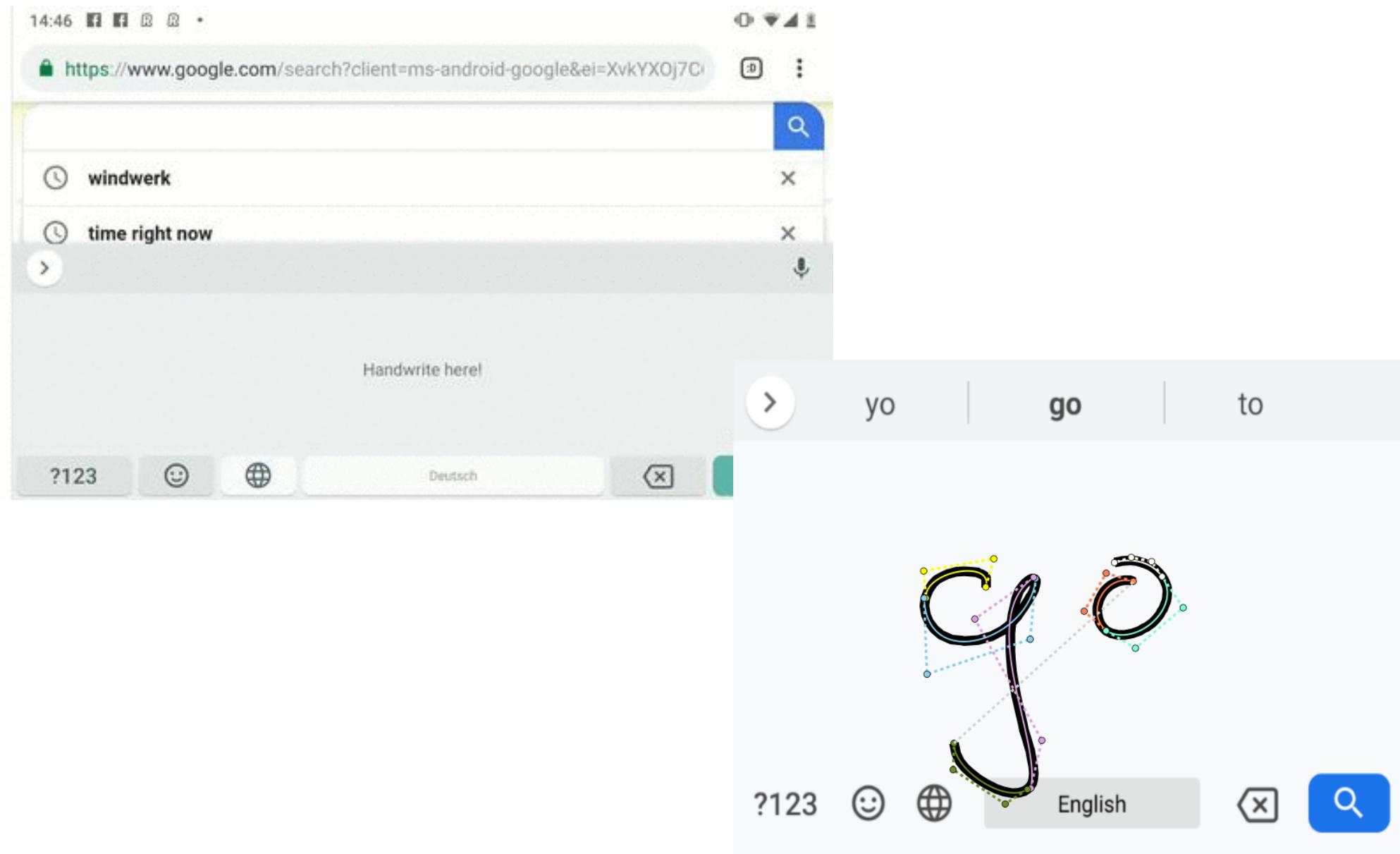


nagenta.tensorflow.org/



GBoard

Guessing what you are writing



Game Over

