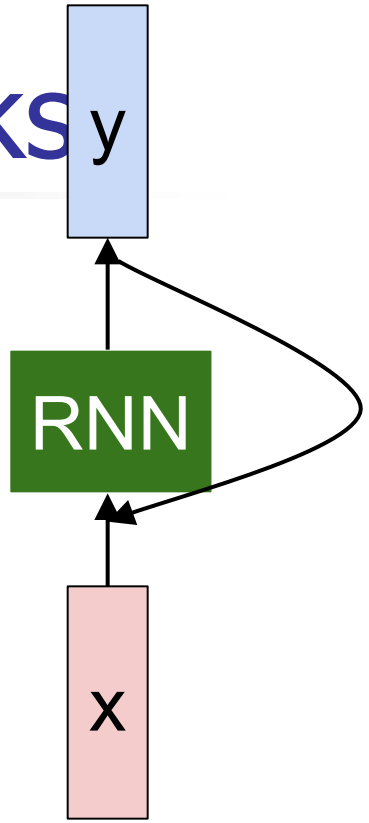
Recurrent Neural Network

Jayanta Mukhopadhyay
Dept. of Computer Science and Engg.

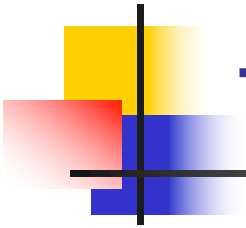
Courtesy: K. Sairam

Recurrent Neural Networks

- Conventional NN (including CNN):
  - Feed forward, No feedback.
  - Fixed size input and output.
  - Fixed number of layers
- Recurrent Neural Networks
  - Feedback from output of a neuron.
  - Sequences in the input, the output, or in the most general case both.
  - Equivalent to unfolded feedforward network of infinite number of layers.



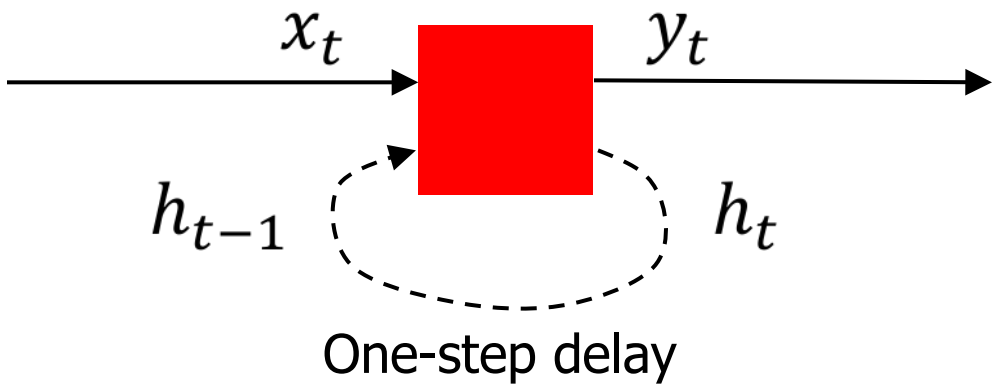
Recurrent Neural Networks: A few Applications



- Semantic labelling of a sequence
  - Classification with the input as a sequence.
- Prediction in a sequence
  - using previous video frames to inform the understanding of the present frame.
  - a language model tries to predict the next word based on the previous ones.
- Sequence translation / generation
  - Input a sequence and output a sequence.

Recurrent Neural Networks

Recurrent networks introduce cycles and a notion of time.

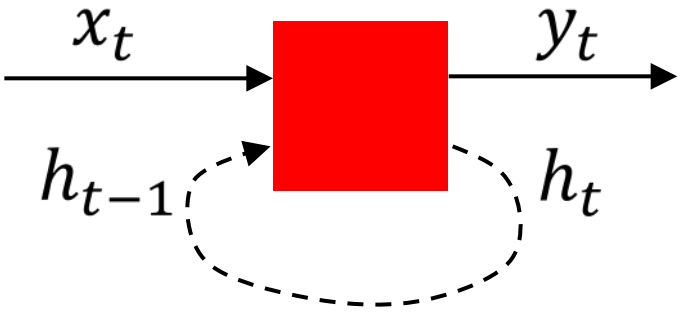


- They are designed to process sequences of data $x_1, \ldots, x_n$ and can produce sequences of outputs $y_1, \ldots, y_m$.

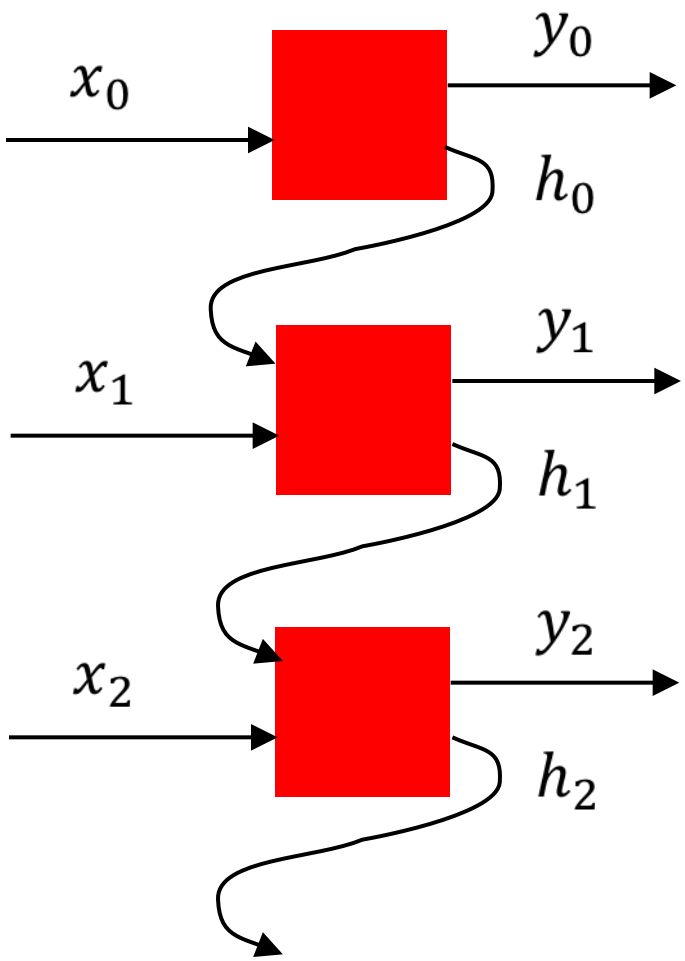
Courtesy: John Canny, UC Berkley.

Unrolling of RNN

Number of stages in unrolling depends on the input sequence length.

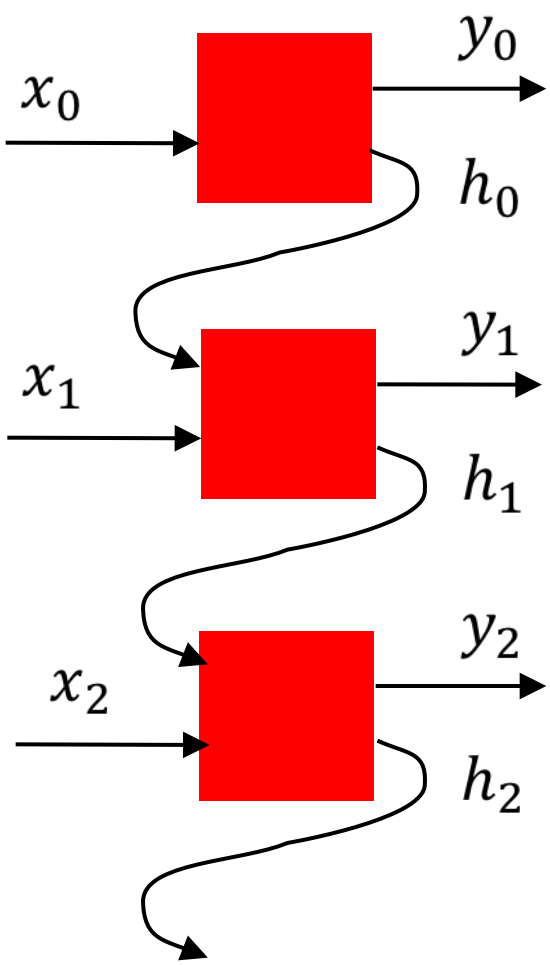


One-step delay

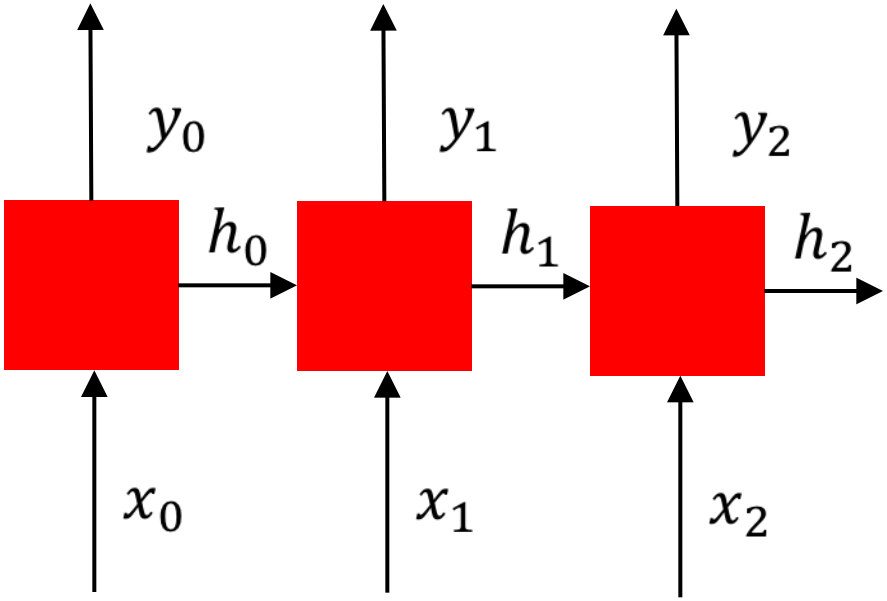


Courtesy: John Canny, UC Berkley.

Unrolling of RNN



Usually drawn as:



Learning algorithm: Back propagation through time (BPTT)

Courtesy: John Canny, UC Berkley.

Recurrent Neural Networks

- usually want to predict a vector at some time steps
- Process a sequence of vectors **x** by applying a recurrence formula at every time step.

$$h_t = f_W(h_{t-1}, x_t)$$

new state

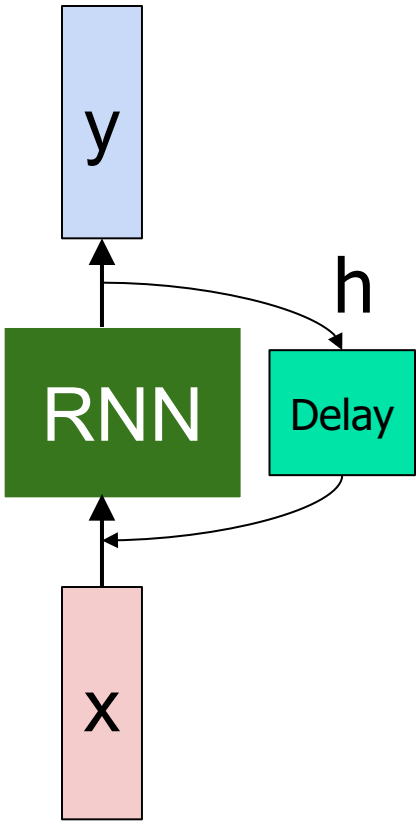
old state

input vector at some time step

some function with parameters W

$$y_t = W_{hy} h_t$$

$$y_t = W_{hy} h_t$$



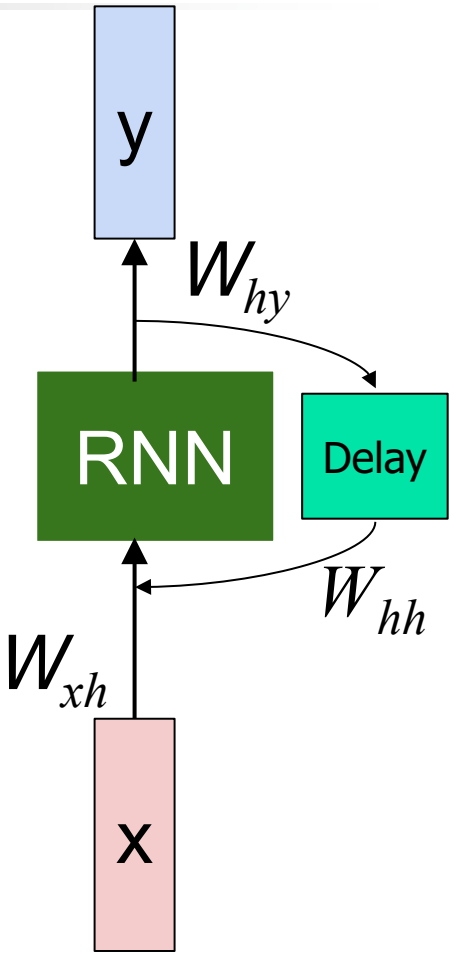
The same function and the same set of parameters are used at every time step.

Recurrent Neural Networks

$$h_t = f_W(h_{t-1}, x_t)$$

$$h_t = \tanh(W_{hh} h_{t-1} + W_{xh} x_t)$$

$$y_t = W_{hy} h_t$$

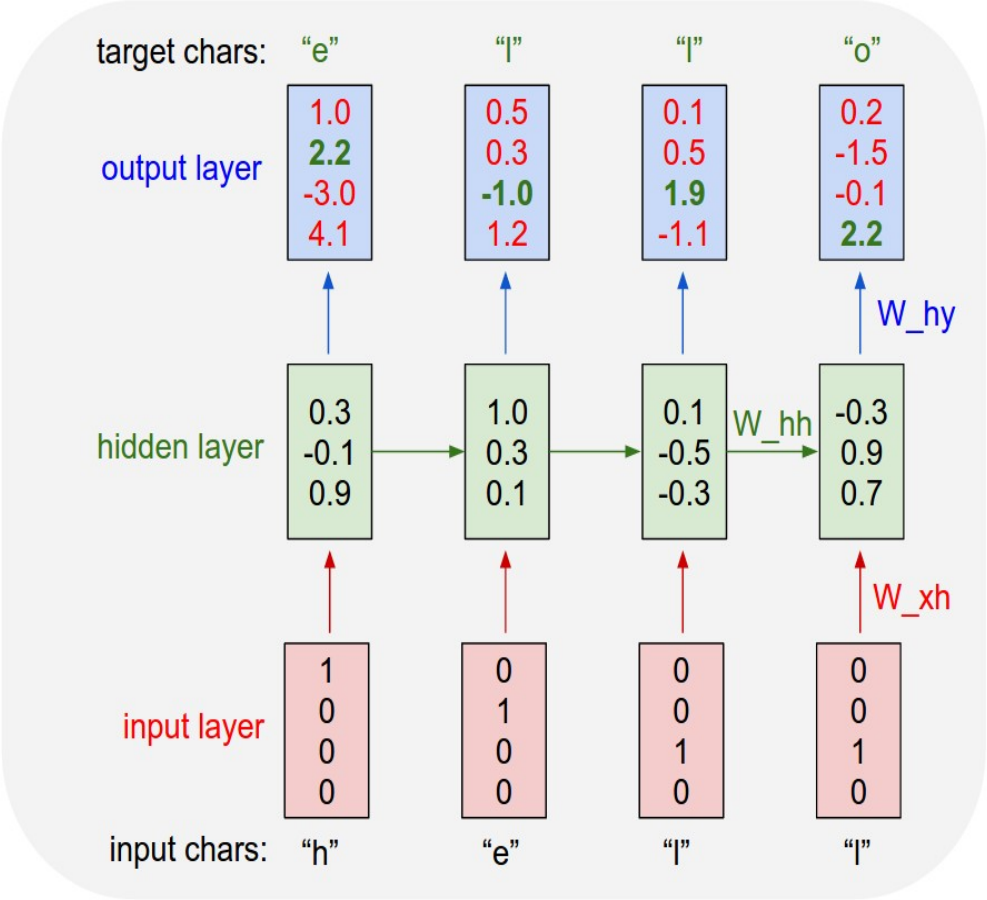


Recurrent Neural Network: An example

Vocabulary: [h,e,l,o]

Example training sequence:
“hello”

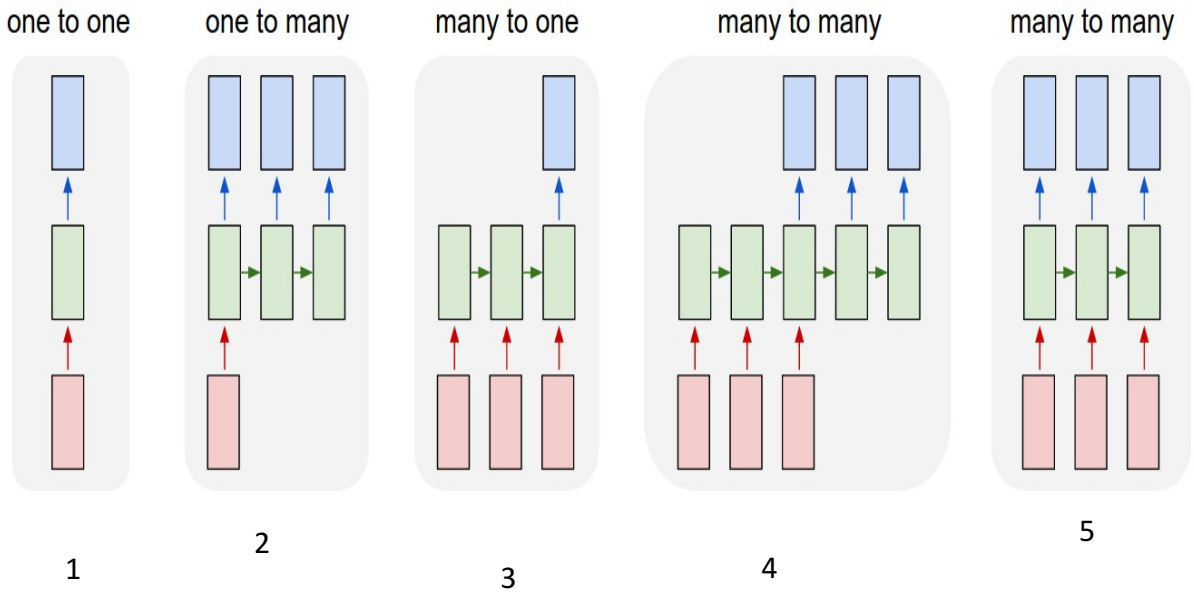
$$h_t = \tanh(W_{hh} h_{t-1} + W_{xh} x_t)$$



Predicting next character of a word.

Courtesy: Andrej Karpathy

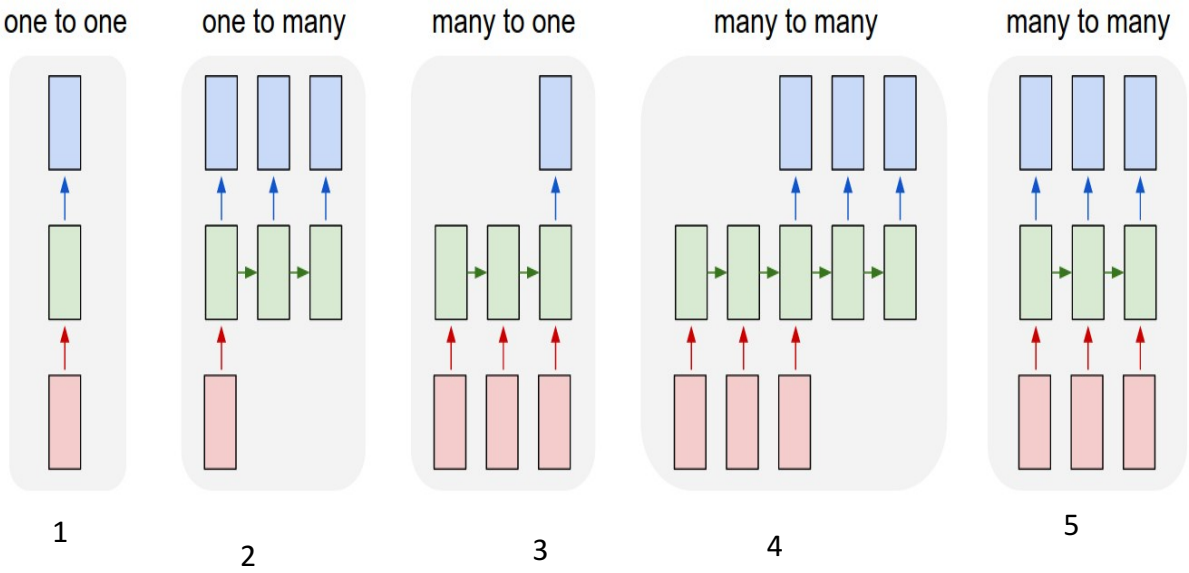
Examples of Recurrent Neural Networks



- Each rectangle is a vector and arrows represent functions (e.g. matrix multiply).
- Input vectors are in red, output vectors are in blue and green vectors hold the RNN's state

Courtesy: Andrej Karpathy

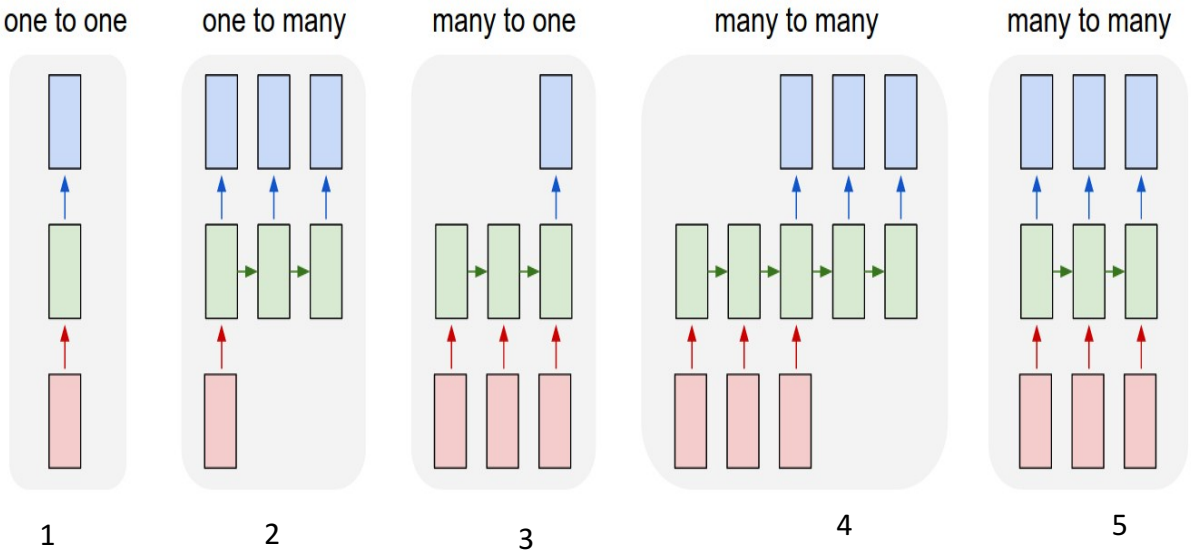
Examples of Recurrent Neural Networks



1. Standard mode of processing without RNN, from fixed-sized input to fixed-sized output (e.g. image classification).
2. Sequence output (e.g. image captioning takes an image and outputs a sentence of words).

Courtesy: Andrej Karpathy

Examples of Recurrent Neural Networks



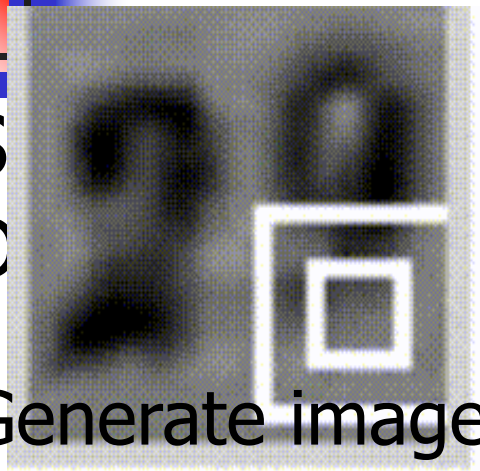
3. Sequence input (e.g. sentiment analysis where a given sentence is classified as expressing positive or negative sentiment).
4. Sequence input and sequence output (e.g. Machine Translation: an RNN reads a sentence in English and then outputs a sentence in French).
5. Synced sequence input and output (e.g. video classification where we wish to label each frame of the video).

Courtesy: Andrej Karpathy

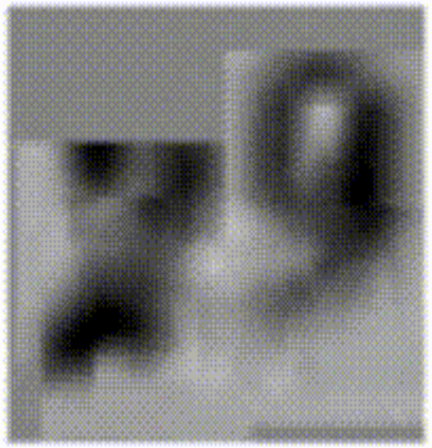
Recurrent Neural Networks

S ocessin

D

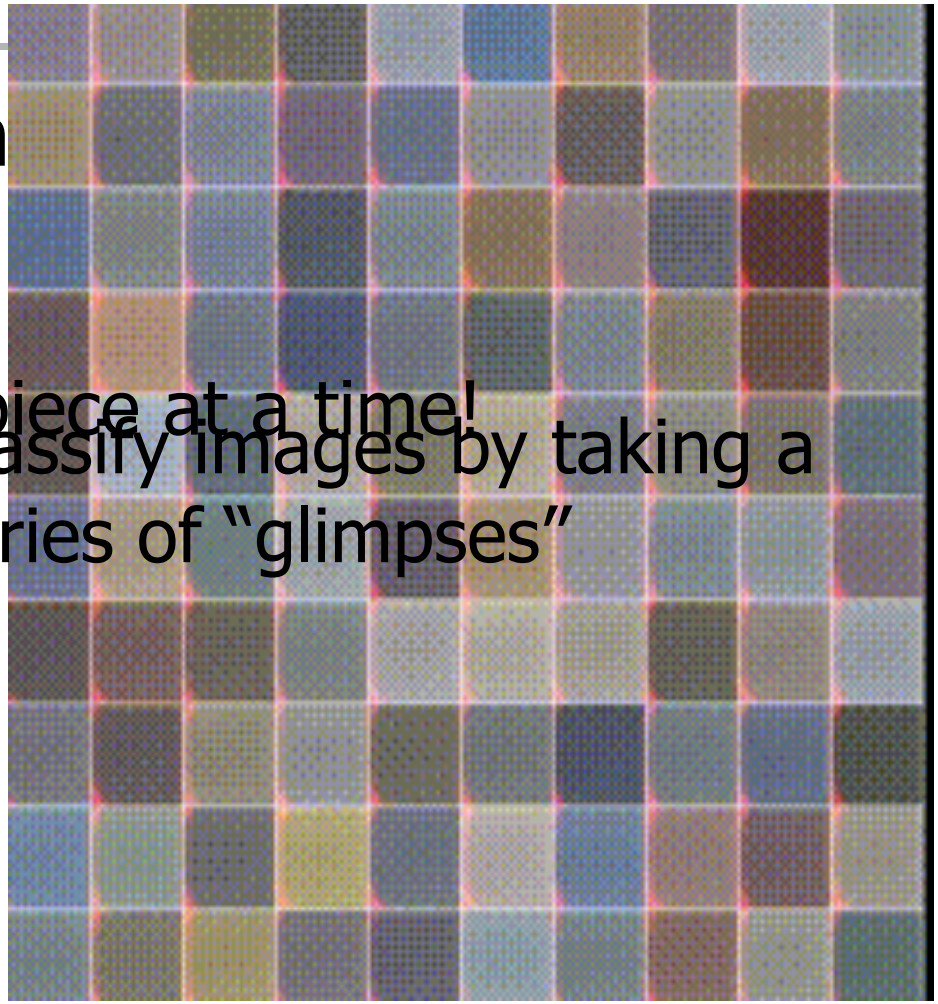


Generate images one piece at a time!



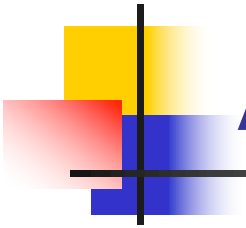
*Multiple Object Recognition with Visual Attention, Ba et al.

Classify images by taking a series of "glimpses"



*DRAW: A Recurrent Neural Network For Image Generation, Gregor et al.

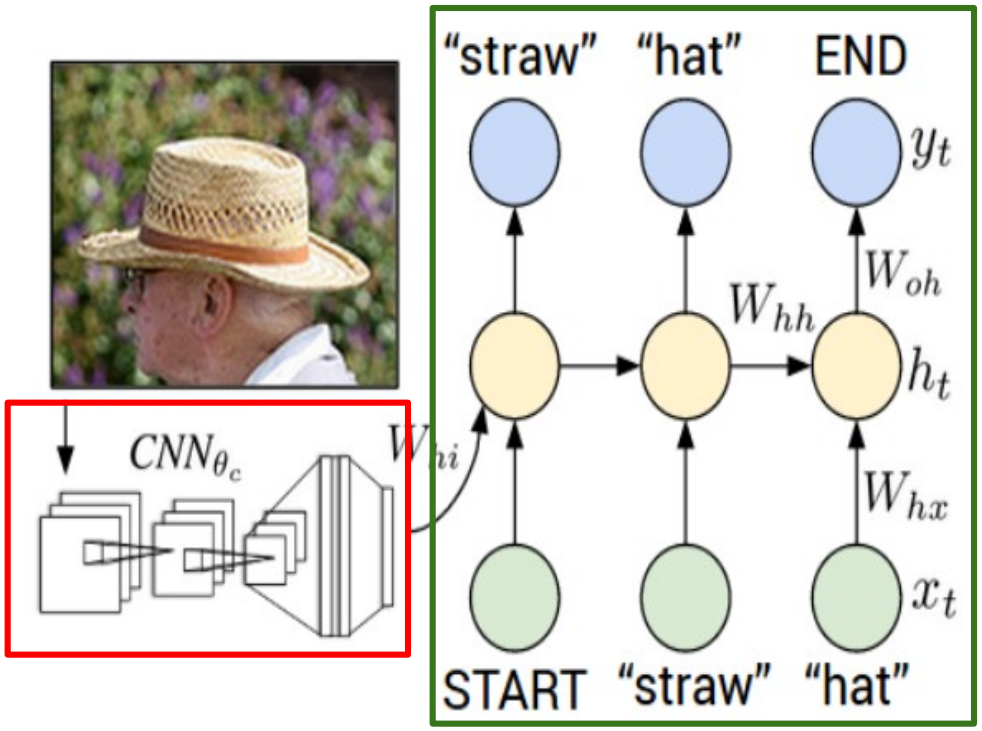
Courtesy: Andrej Karpathy



Applications

1. Image Classification
2. Object Recognition and Localization
3. Image Segmentation
4. Image Captioning

Image Captioning



* Show and Tell: A Neural Image Caption Generator, Vinyals et al.

image

conv-64

conv-64

maxpool

conv-128

conv-128

maxpool

conv-256

conv-256

maxpool

conv-512

conv-512

maxpool

conv-512

conv-512

maxpool

FC-4096

FC-4096

X

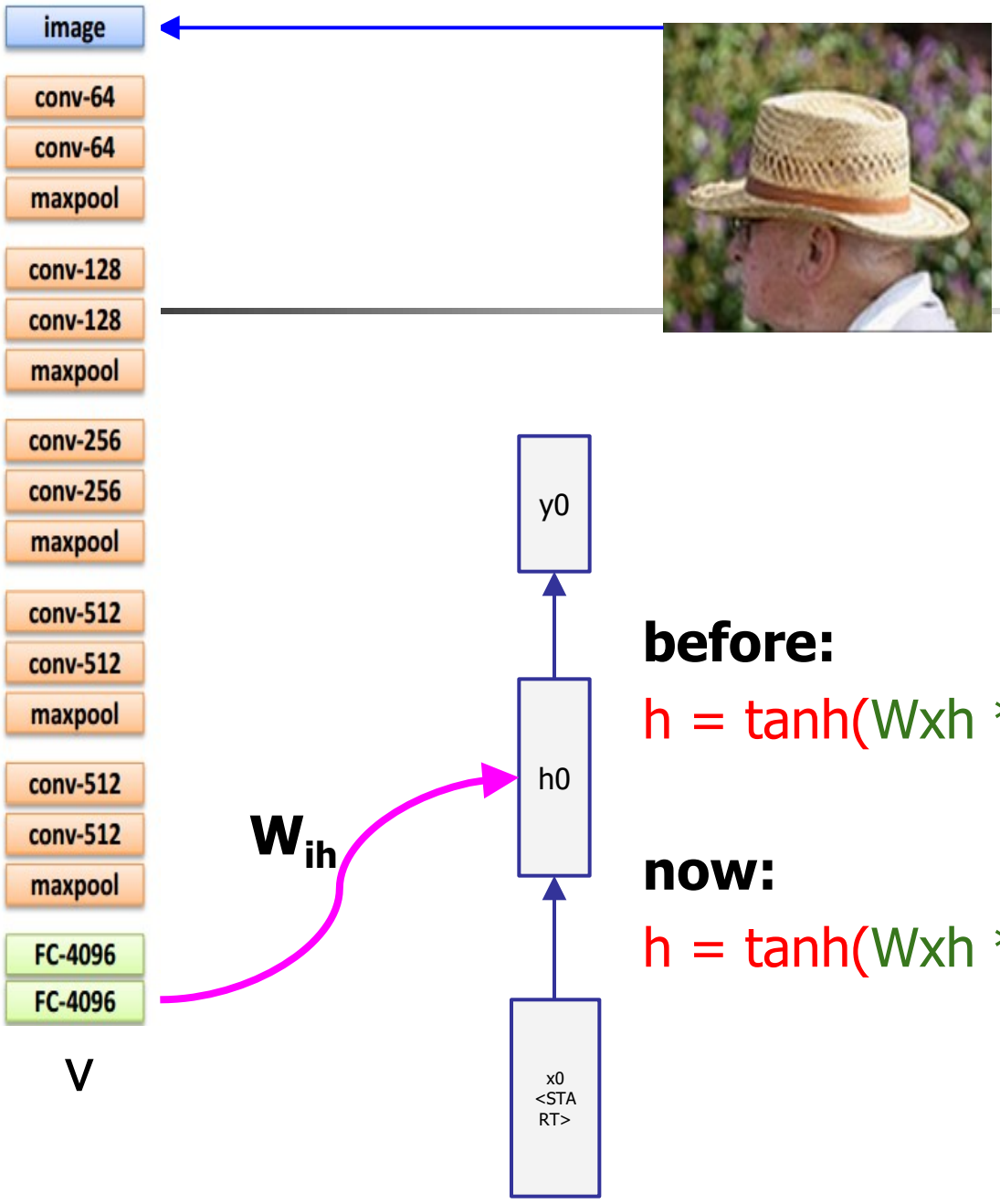


image
conv-64
conv-64
maxpool
conv-128
conv-128
maxpool
conv-256
conv-256
maxpool
conv-512
conv-512
maxpool
conv-512
conv-512
maxpool
FC-4096
FC-4096
V
W_{ih}
y0
h0
x0
<STA
RT>
<START>
before:
h = tanh(Wxh * x + Whh * h)
now:
h = tanh(Wxh * x + Whh * h + Wih * v)

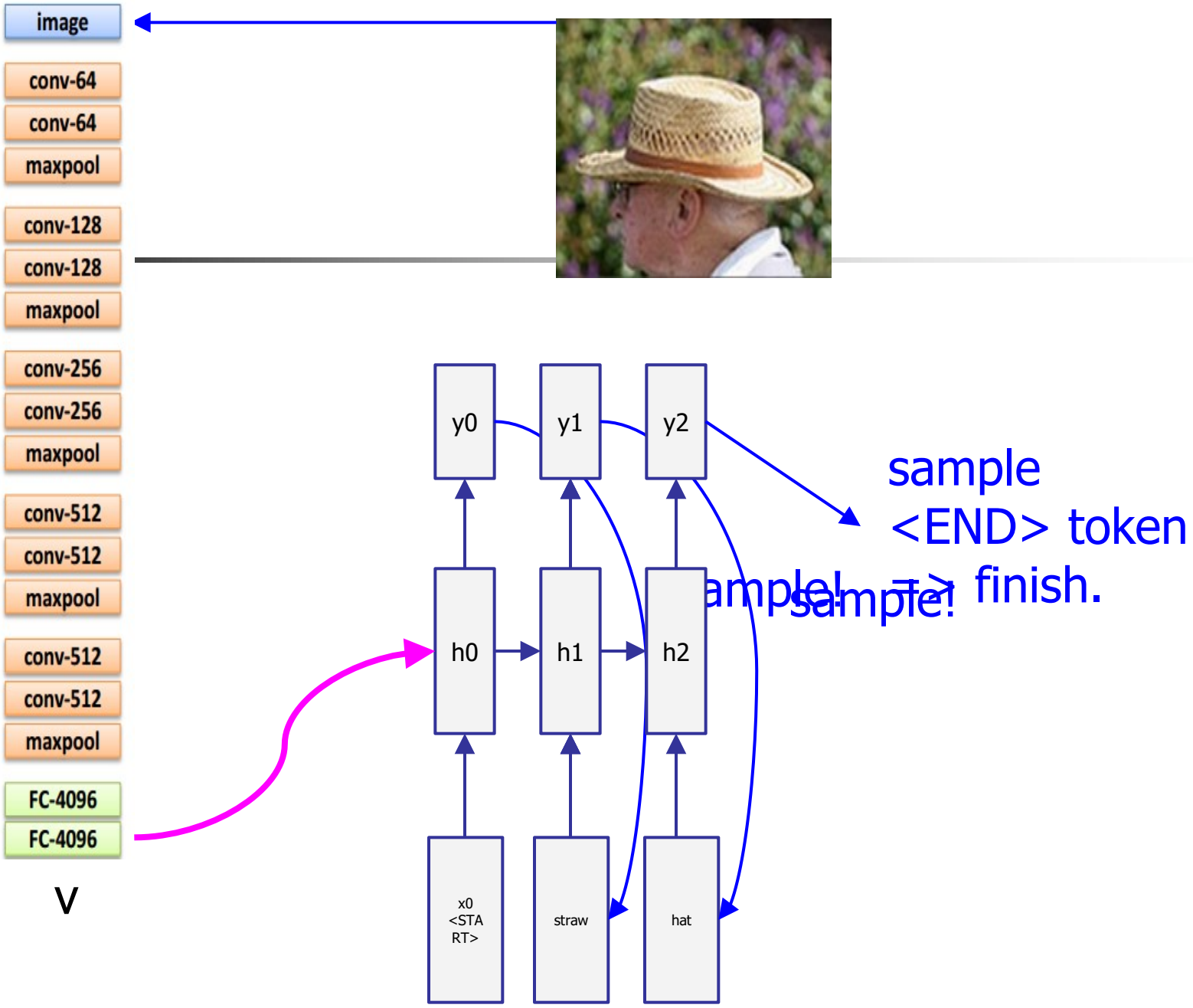


image
conv-64
conv-64
maxpool
conv-128
conv-128
maxpool
conv-256
conv-256
maxpool
conv-512
conv-512
maxpool
conv-512
conv-512
maxpool
FC-4096
FC-4096
v
y0
y1
y2
h0
h1
h2
x0
<STA
RT>
straw
hat
<START>
sample
<END> token
=> finish.
sample!

Image Sentence Datasets

a man riding a bike on a dirt path through a forest.
bicyclist raises his fist as he rides on desert dirt trail.
this dirt bike rider is smiling and raising his fist in triumph.
a man riding a bicycle while pumping his fist in the air.
a mountain biker pumps his fist in celebration.

Microsoft COCO

[Tsung-Yi Lin et al. 2014]

currently:
~120K images
~5 sentences each

"man in black shirt is playing guitar."

"construction worker in orange safety vest is working on road."

"two young girls are playing with lego toy."

"boy is doing backflip on wakeboard."

"a young boy is holding a baseball bat."

"a cat is sitting on a couch with a remote control."

"a woman holding a teddy bear in front of a mirror."

"a horse is standing in the middle of a road."


Thank you!