Today: Outline

- Feature Extraction and Transfer Learning
- RNNs

- Reminders: *Please be sure to record in-class bonuses*

Pre-lec Material 2,
due: Friday, Jun 4

Problem Set 1,
due: Friday, Jun 4

Grades Web Page

- http://cs-people.bu.edu/sbargal/cs523/grades.html

Deep Learning

CS 523 Summer1 2021

You can use your random ID to access your row of grades for the course.

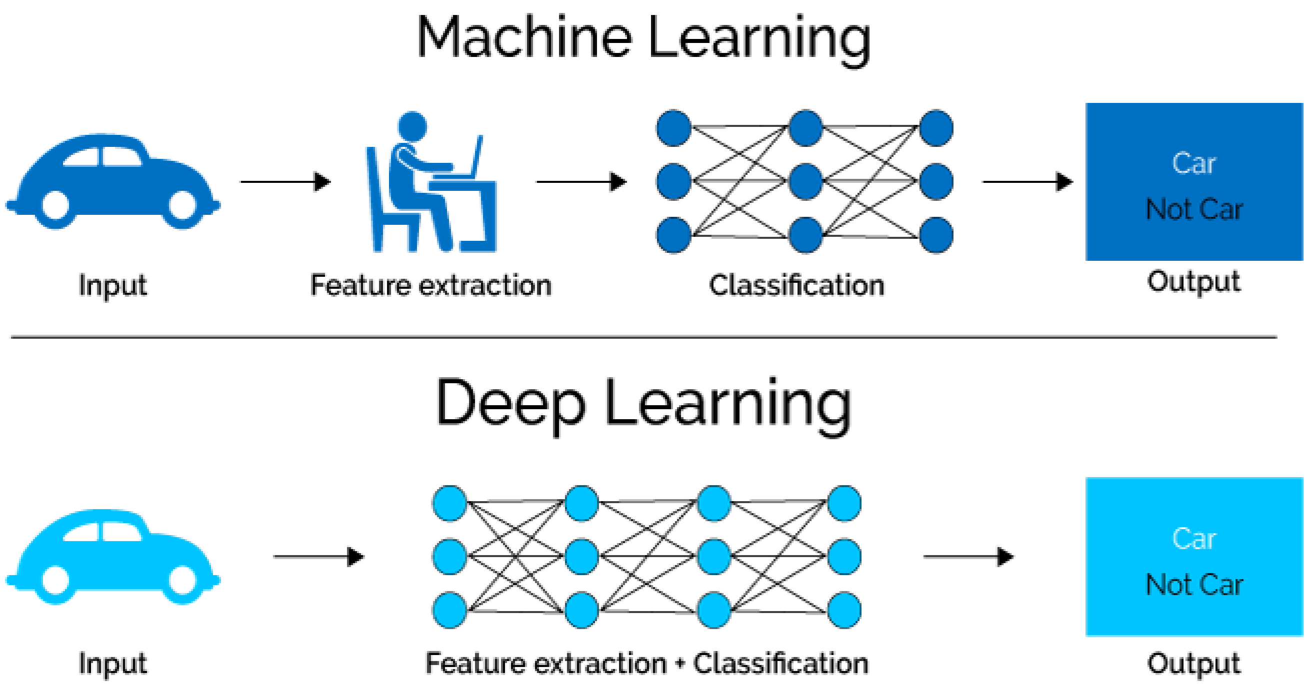
Grades**

Rand ID	PS1 (/100)	PS2 (/100)	PS3 (/100)	Par (/5)	Bonus (/2)	Pre-lec (/1)	Exam (/100)	Project (/100)
42	-	-	-	-	-	1	-	-
29	-	-	-	-	-	1	-	-
41	-	-	-	-	-	1	-	-
14	-	-	-	-	-	0.75	-	-
47	-	-	-	-	-	1	-	-
18	-	-	-	-	-	0	-	-
37	-	-	-	-	-	1	-	-
46	-	-	-	-	-	1	-	-
23	-	-	-	-	-	1	-	-
48	-	-	-	-	-	1	-	-
44	-	-	-	-	-	1	-	-
45	-	-	-	-	-	1	-	-
31	-	-	-	-	-	0.5	-	-
27	-	-	-	-	-	1	-	-

Neural Networks

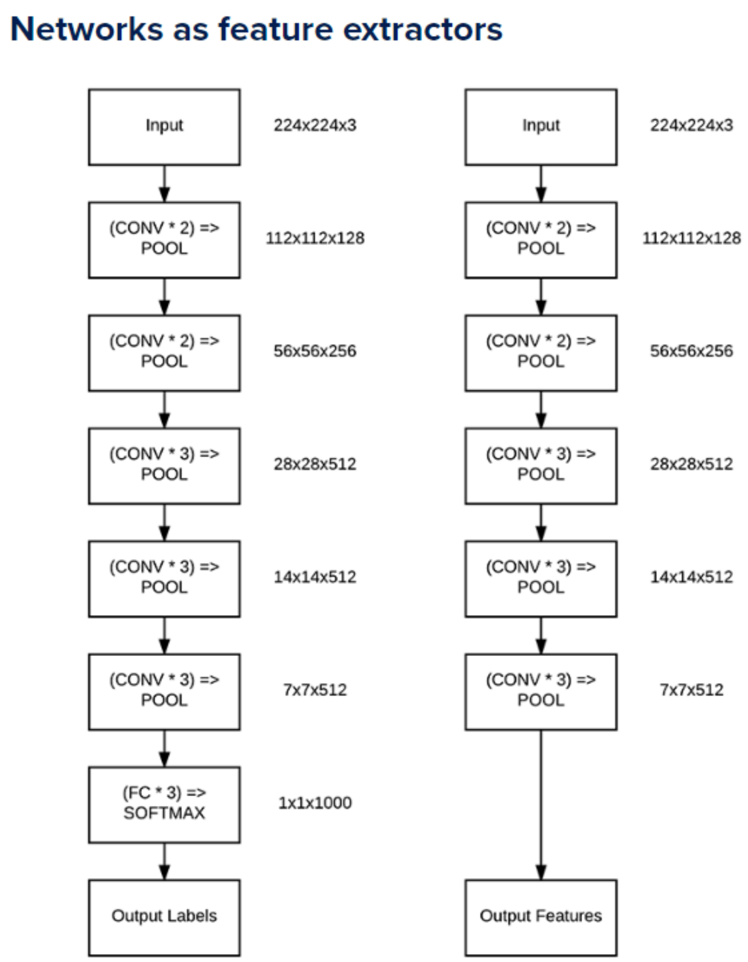
Feature Extraction and Transfer Learning

Feature Extraction



[Towards Data Science]

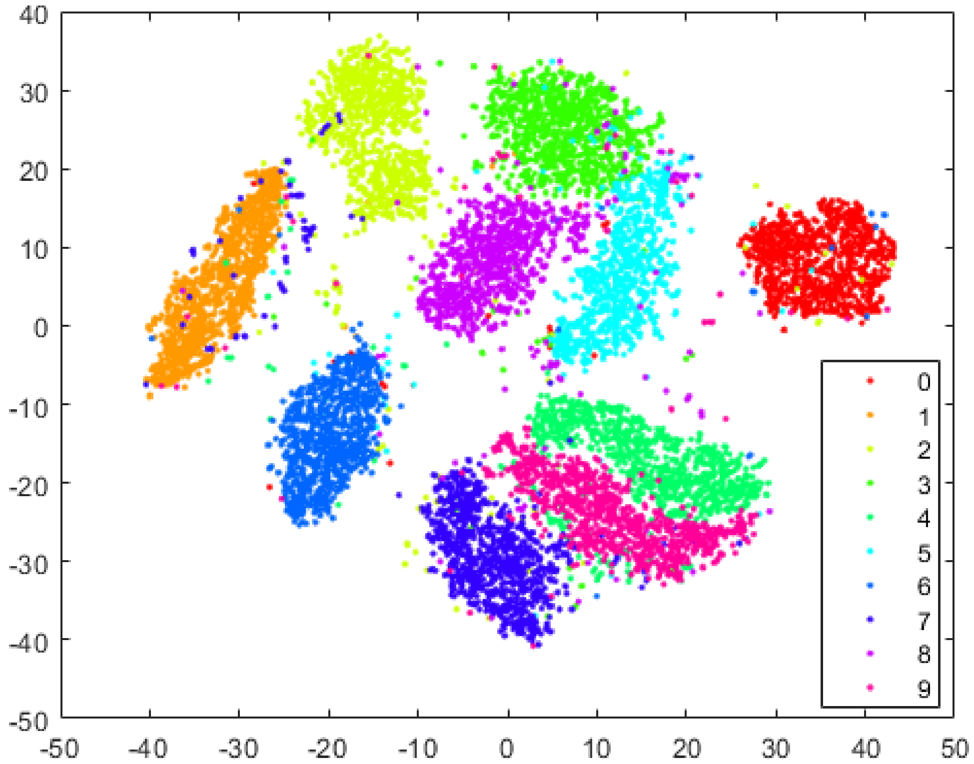
Feature Extraction From NNs



[Towards Data Science]

Feature Visualization

- T-SNE
  A common way for visualizing features through dimensionality reduction.



Pre-training

- Simply put, a **pre-trained model** is a **model** created by some one else to solve a similar problem.

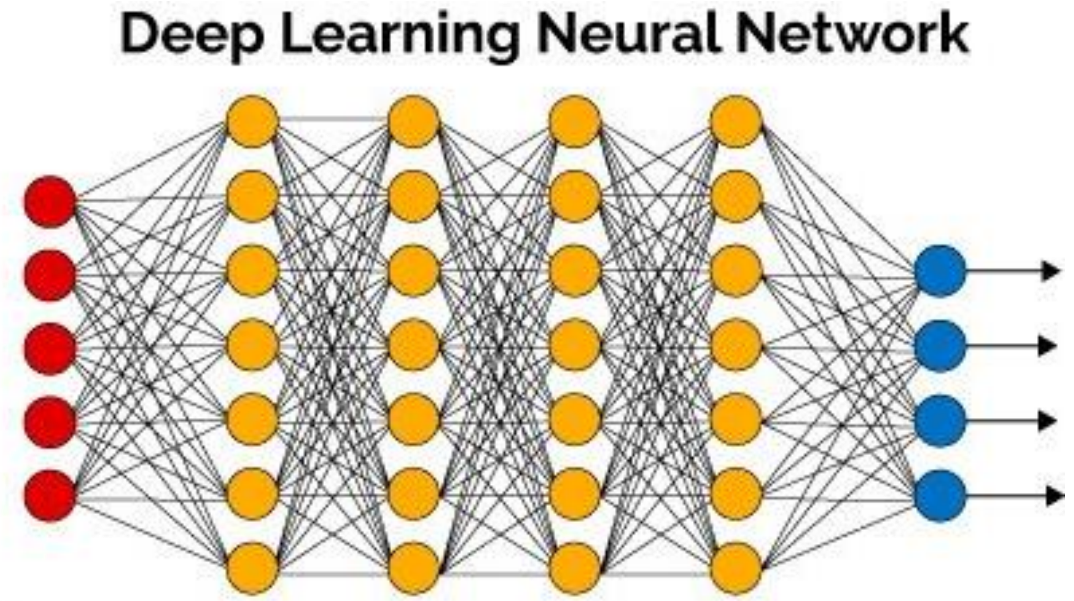
- Instead of building a **model** from scratch to solve a similar problem, you use the **model trained** on other problem as a starting point.

- This catalyzed academic research significantly.

- Model Zoo.

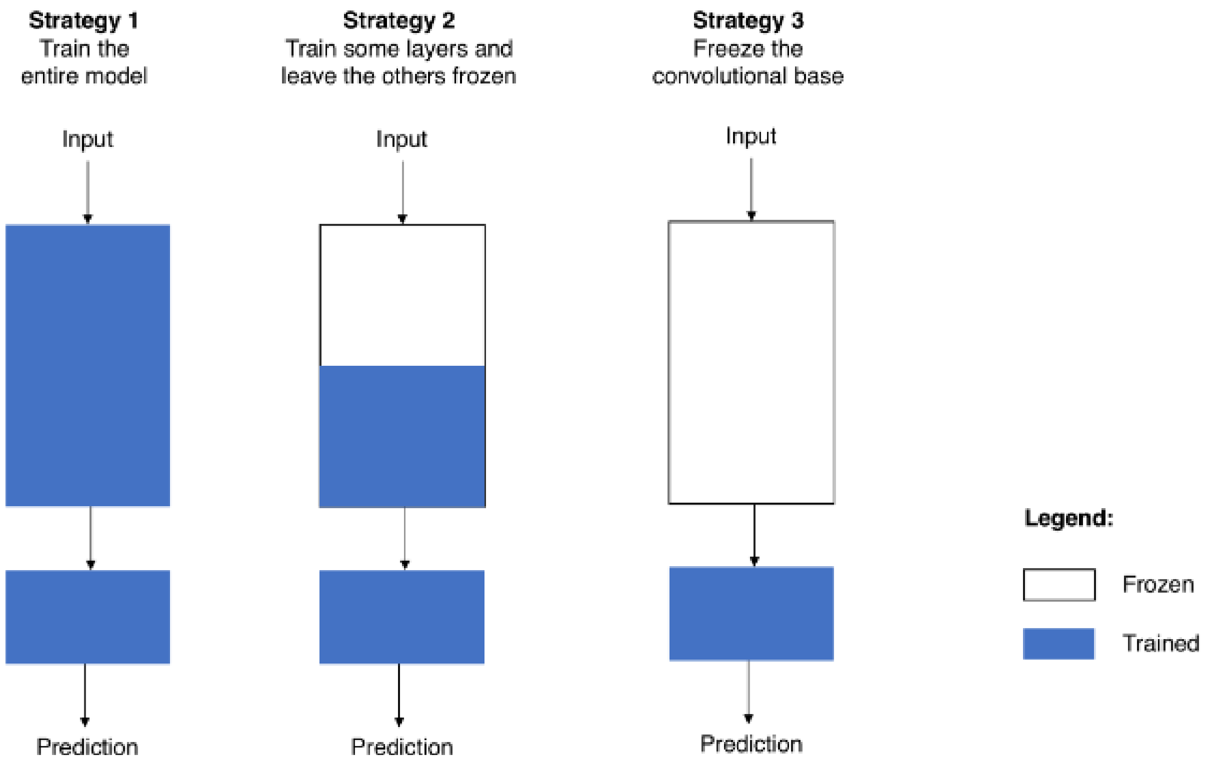
Fine-tuning

- What if the number of output neurons (or classes) is not the same?



https://heartbeat.fritz.ai/pre-trained-machine-learning-models-vs-models-trained-from-scratch-63e079ed648f

Transfer Learning — the Pre-train & Fine-tune Paradigm



https://heartbeat.fritz.ai/pre-trained-machine-learning-models-vs-models-trained-from-scratch-63e079ed648f

Modeling Sequences - RNNs

Limitations of Feed-Fwd Networks

- Limitations of feed-forward networks
  - **Fixed length**
  Inputs and outputs are of fixed lengths
  - **Independence**
  Data (example: images) are independent of one another

Advantages of RNN Models

- What feed-forward networks cannot do

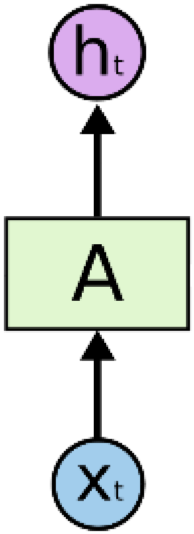
  - **Variable length**

    "We would like to accommodate temporal sequences of various lengths."

  - **Temporal dependence**

    "To predict where a pedestrian is at the next point in time, this depends on where he/she were in the previous time step(s)."

Vanilla Neural Network (NN)

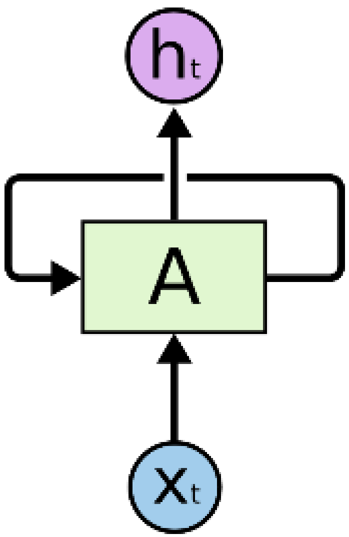


- NN

x_t: input/event
h_t: output/prediction
A : chunk of NN

Every input is treated independently.

Recurrent Neural Network (RNN)



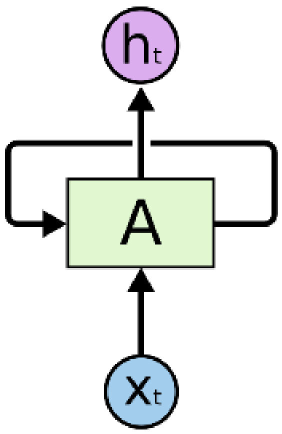
- RNN

The loop allows information to be passed from one time step to the next.

Now we are modeling the dynamics.

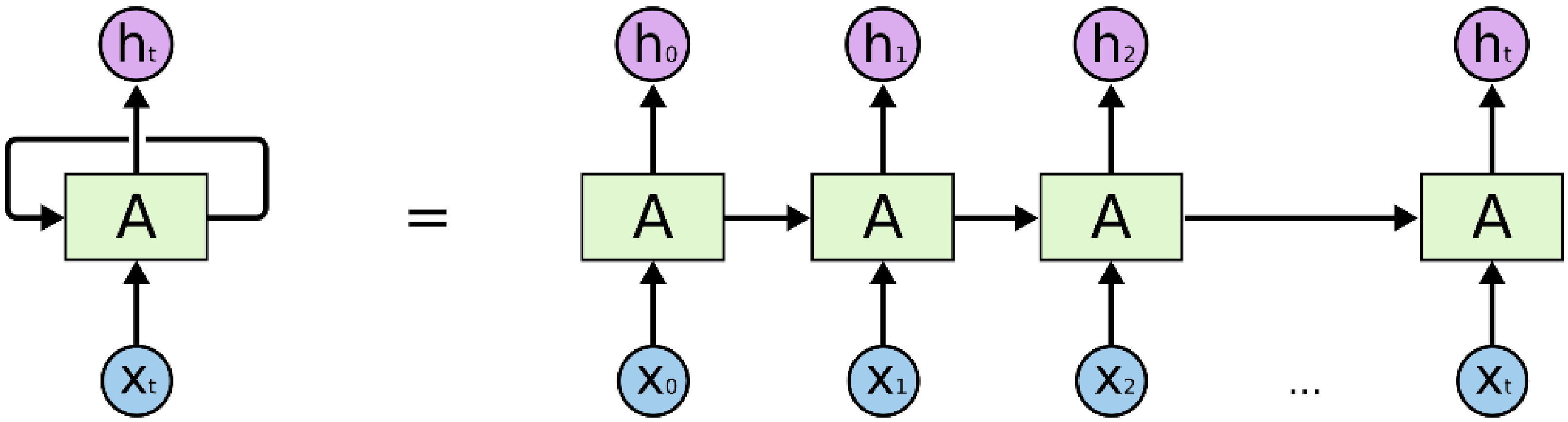
Recurrent Neural Network (RNN)

- A recurrent neural network can be thought of as multiple copies of the same network, each passing a message to a successor.

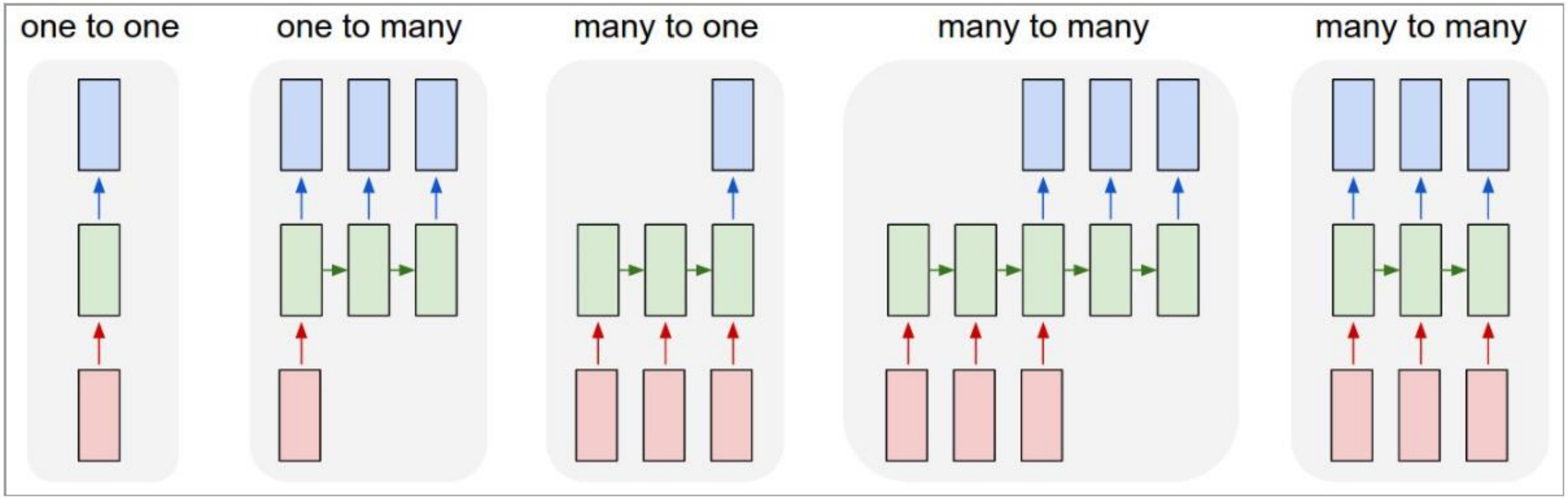


Recurrent Neural Network (RNN)

- A recurrent neural network can be thought of as multiple copies of the same network, each passing a message to a successor.



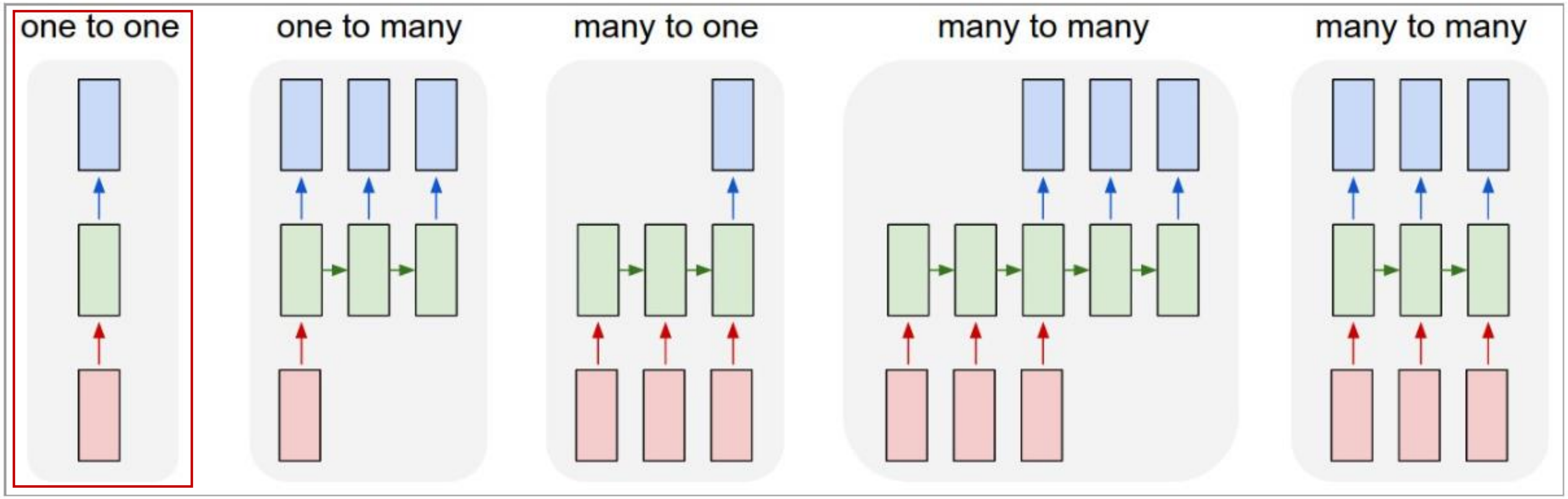
RNN Architectures



Vanilla NN

Recurrent NNs

One-to-one



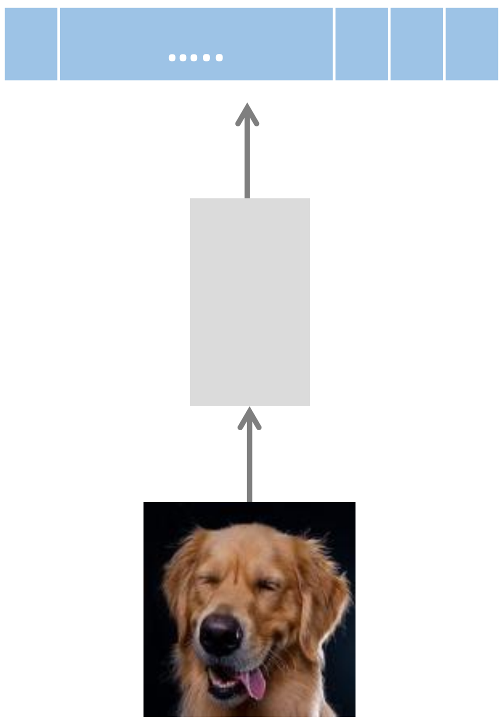
Vanilla mode of processing without RNN

Example: Image classification

Example: One-to-one

Vanilla mode of processing without RNN

Example: Image classification

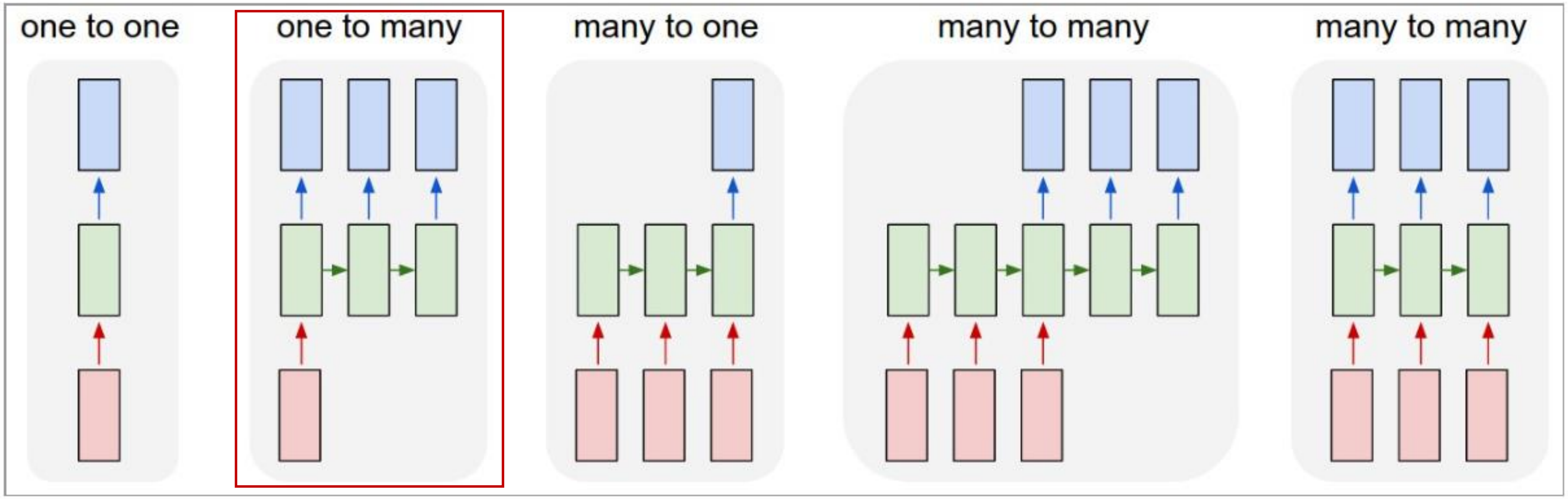


Prediction: "Dog"

Vanilla NN

Image

One-to-many



Sequence output

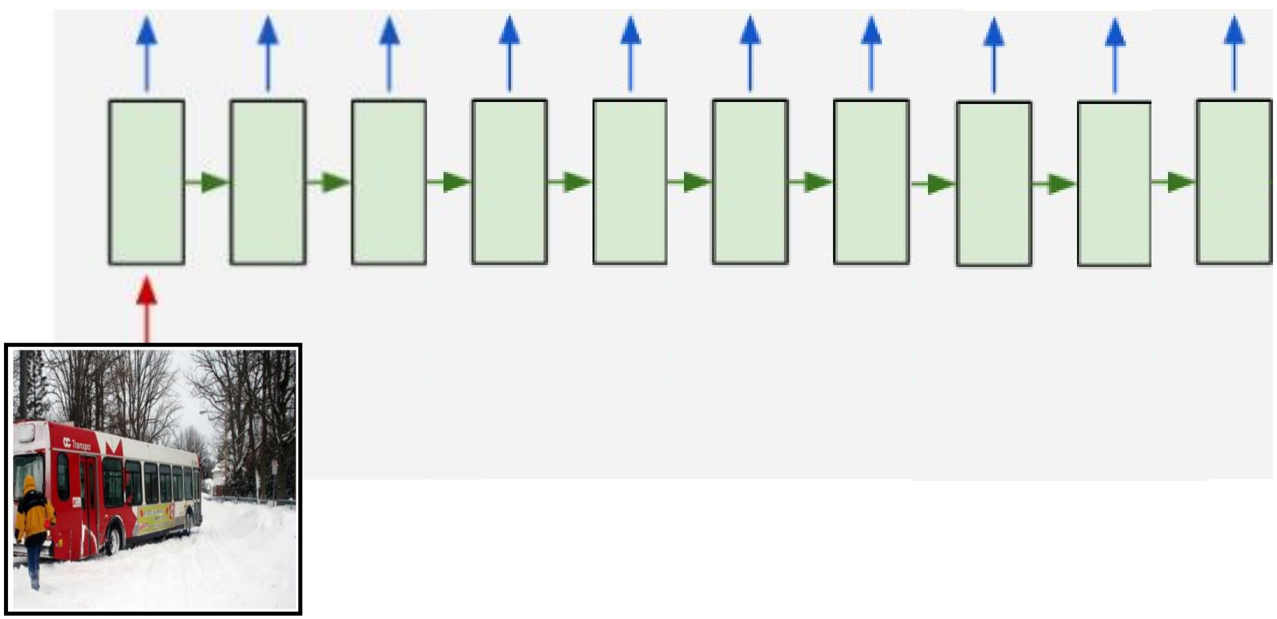
Example: Image captioning

Example: One-to-many

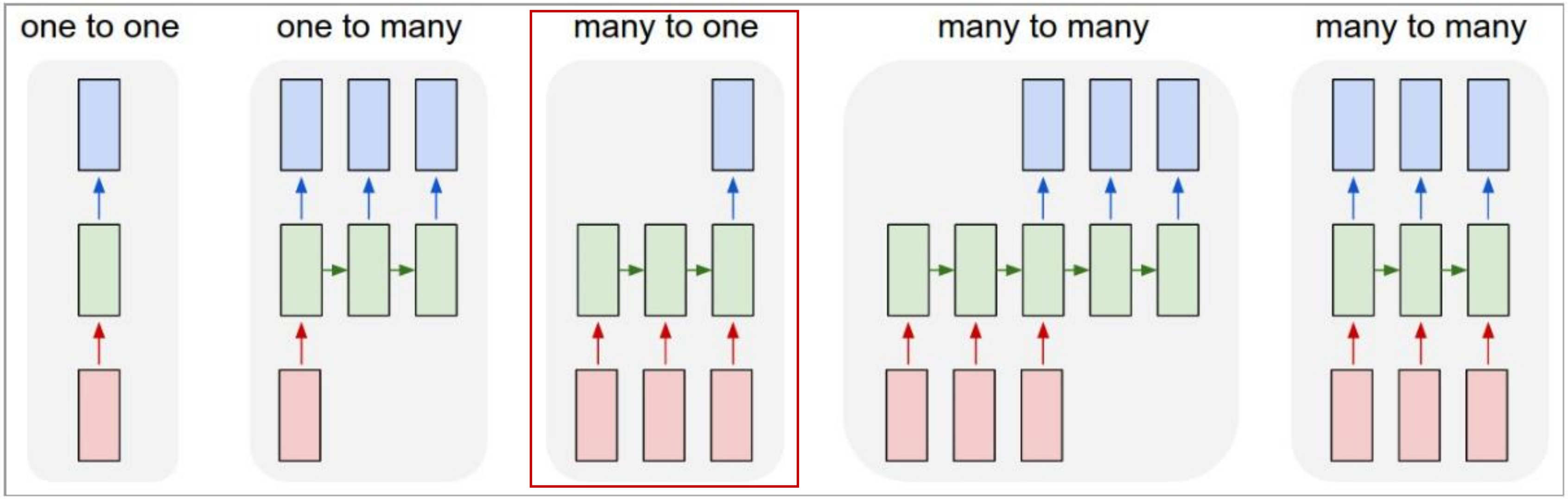
Sequence output

Example: Image Captioning

Bus driving down a snowy road next to trees <EOS>



Many-to-one



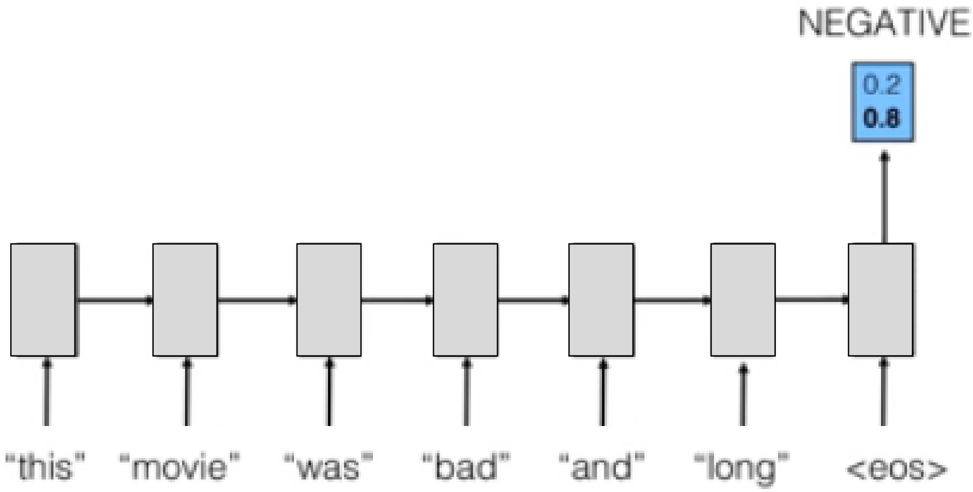
Sequence input

Examples: Sentiment analysis
Action recognition

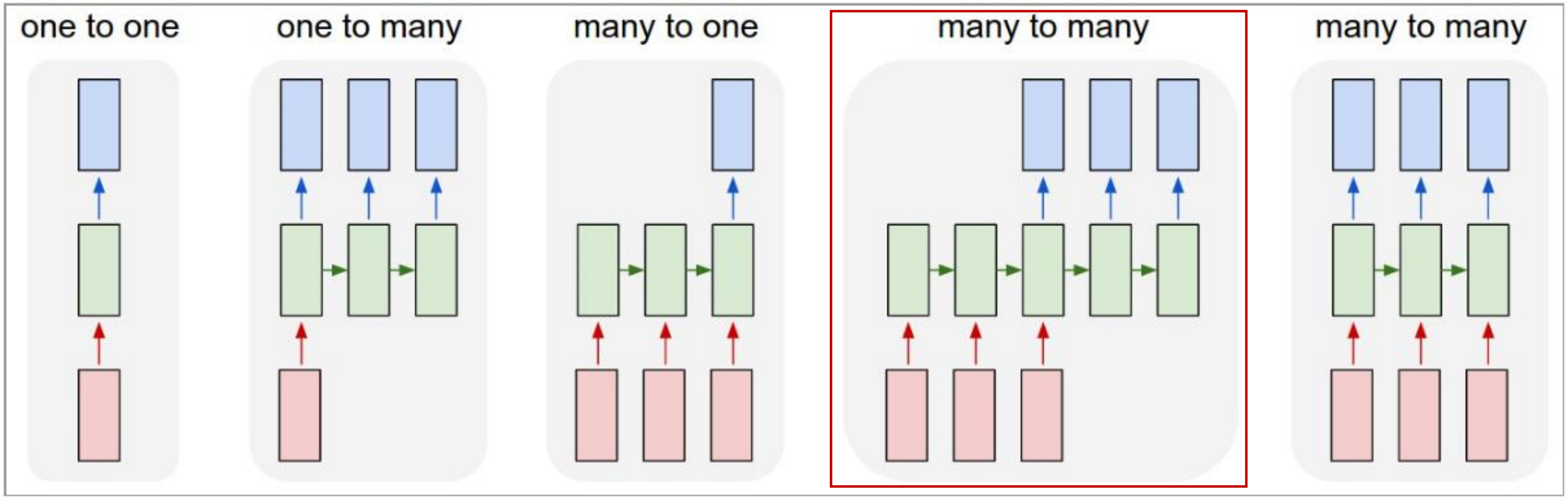
Example: Many-to-one

Sequence input

Example: Sentiment analysis



Many-to-many



Sequence input and sequence output

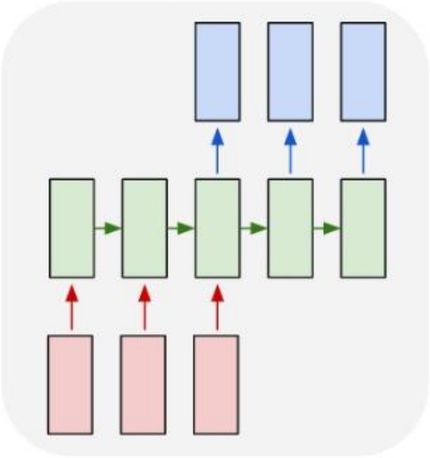
Example: Machine translation

Example: Many-to-many

Sequence input and sequence output

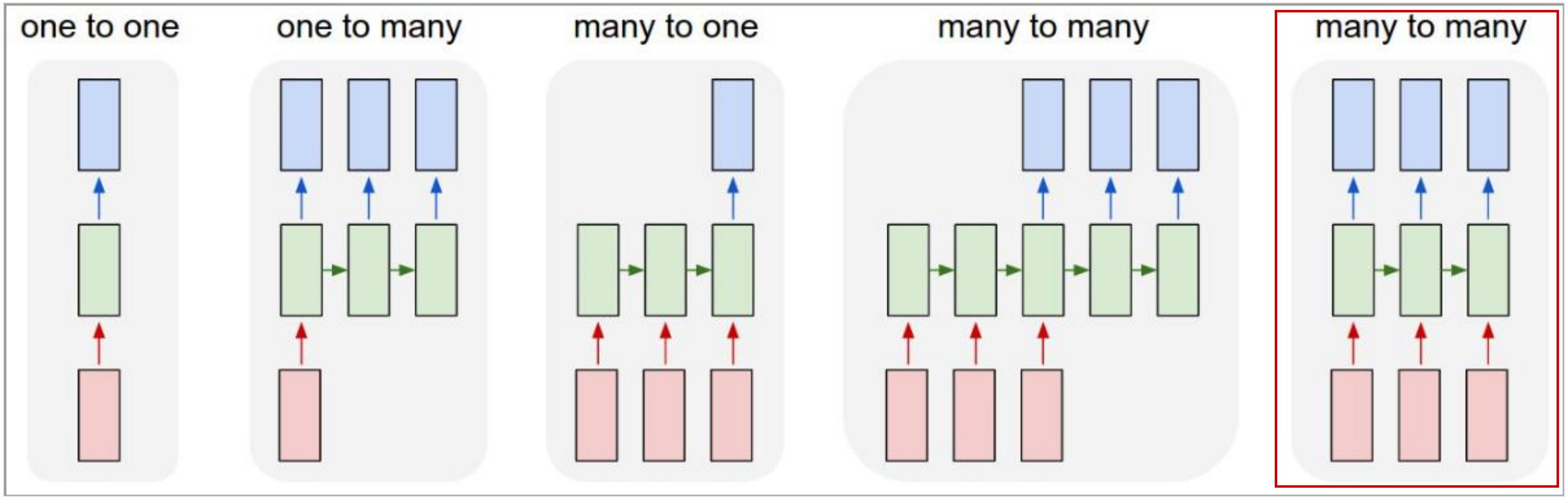
Example: Machine translation

French Translation



English Sentence

Synced Many-to-many

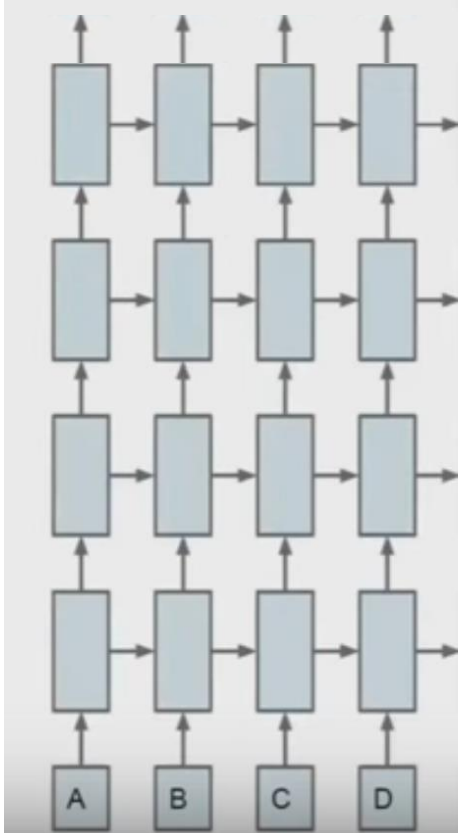


Synced sequence input and output

Examples: Tracking

Early action detection

Deep RNNs



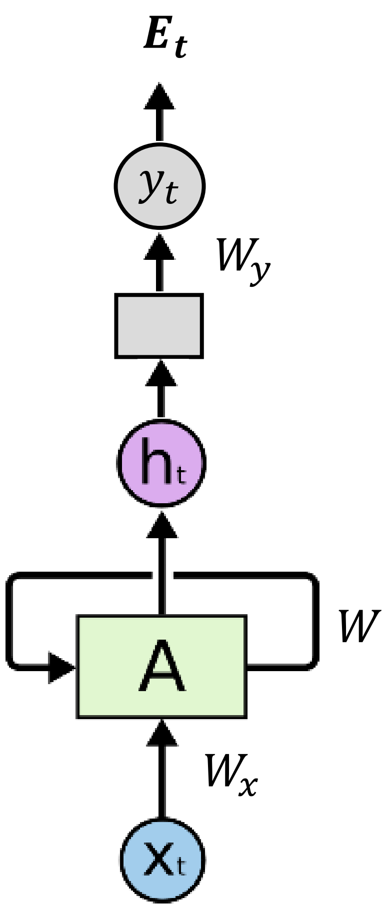
- Stacking RNNs
- More ways of propagating information!
- Requires a lot of data!

Fwd RNN TT

- Forward pass through time

$$h_t = W\phi(h_{t-1}) + W_x x_t$$

$$y_t = W_y \phi(h_t)$$

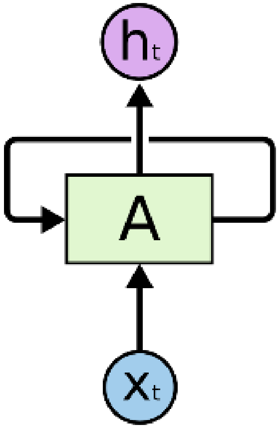


Recurrent Neural Network (RNN)

Aside: Forward pass

$$h_t = W\phi(h_{t-1}) + W_x x_t$$

$$y_t = W_y \phi(h_t)$$



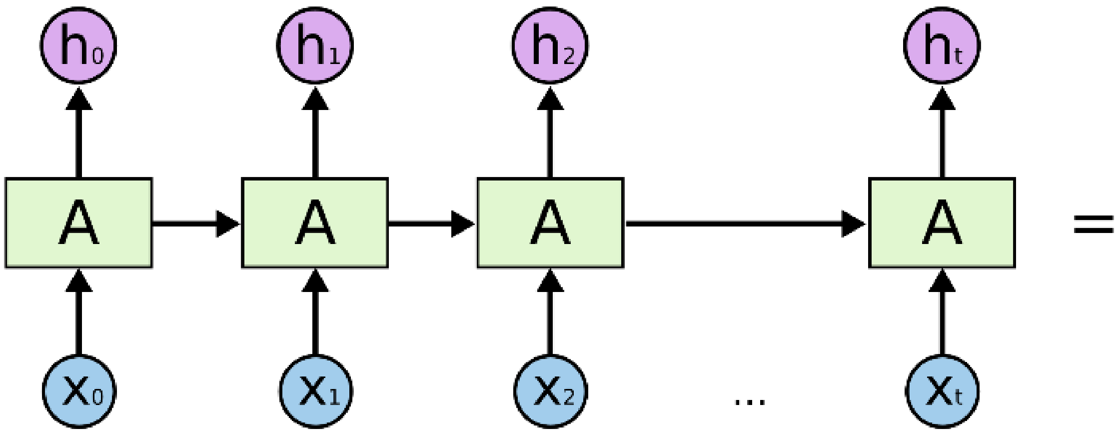
Recurrent Neural Network (RNN)

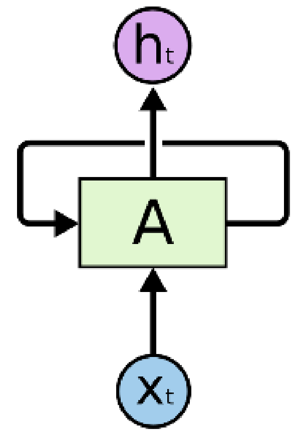
- Error or cost is computed for each prediction.

Aside: Forward pass

$$h_t = W\phi(h_{t-1}) + W_x x_t$$

$$y_t = W_y \phi(h_t)$$



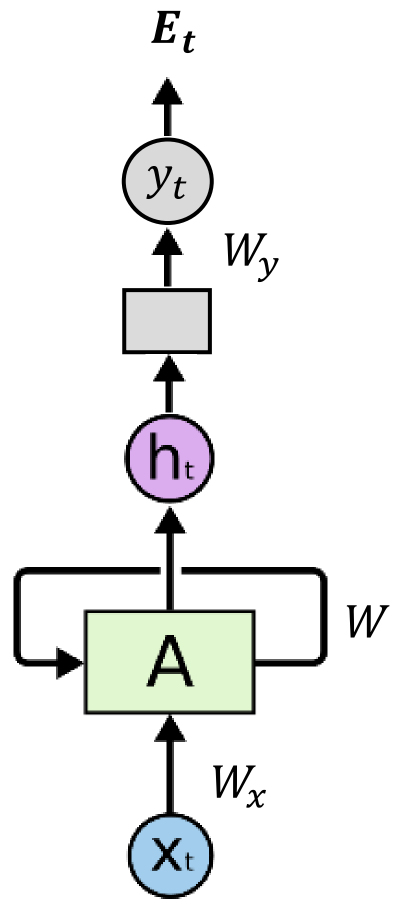


BP TT

- Backpropagation through time

$$\frac{\partial E}{\partial W} = \sum_{t=1}^{T} \frac{\partial E_t}{\partial W}$$

Aside: Forward pass



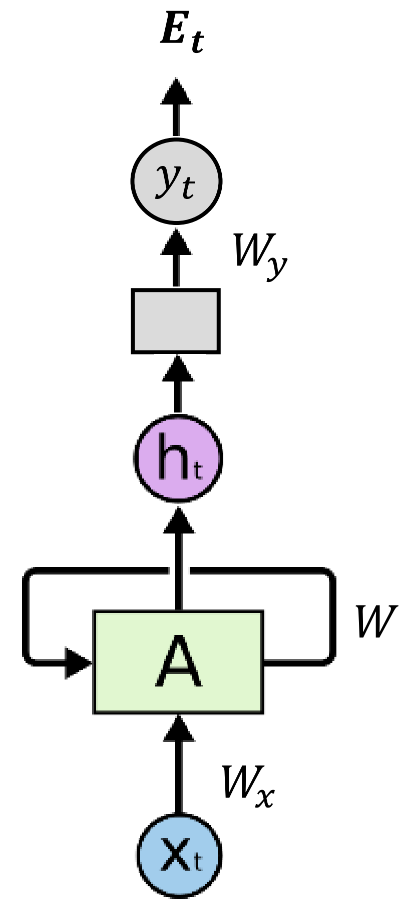
BP TT

- Backpropagation through time

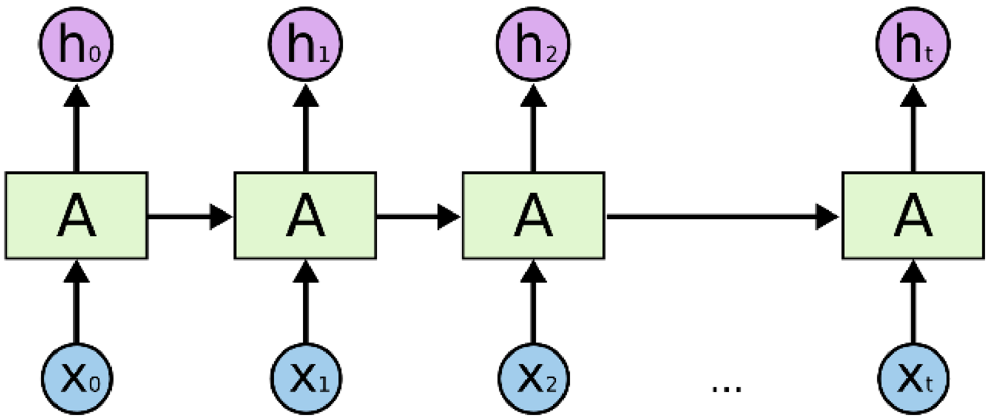
$$\frac{\partial E}{\partial W} = \sum_{t=1}^{T} \boxed{\frac{\partial E_t}{\partial W}}$$

$$\frac{\partial E_t}{\partial W} = \sum_{k=1}^{t} \frac{\partial E_t}{\partial y_t} \frac{\partial y_t}{\partial h_t} \frac{\partial h_t}{\partial h_k} \frac{\partial h_k}{\partial W}$$

Aside: Forward pass

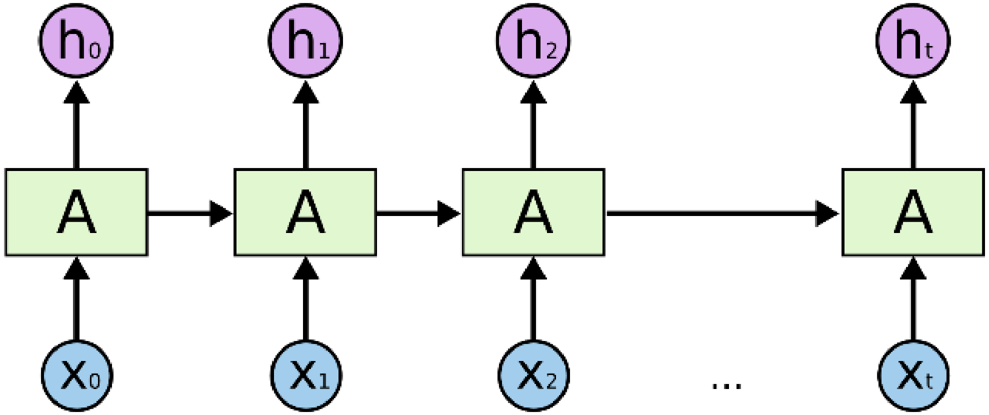


BP TT



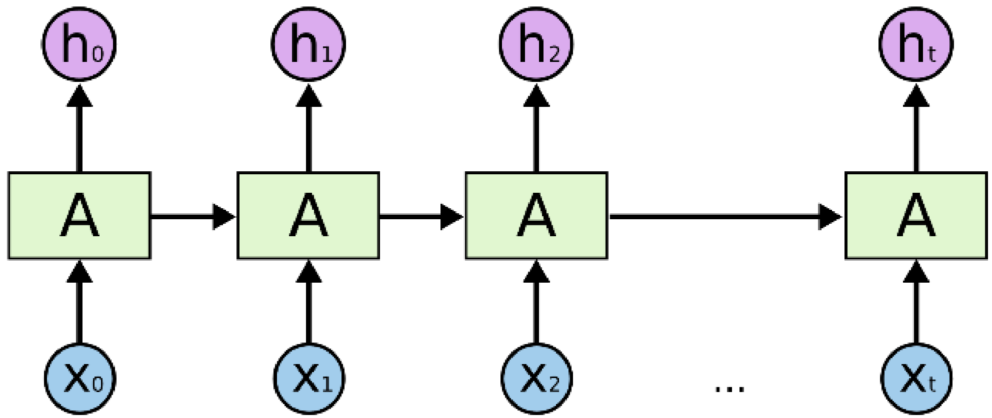
$$\frac{\partial E_t}{\partial W} = \sum_{k=1}^{t} \frac{\partial E_t}{\partial y_t} \frac{\partial y_t}{\partial h_t} \frac{\partial h_t}{\partial h_k} \frac{\partial h_k}{\partial W}$$

BP TT



$$\frac{\partial E_t}{\partial W} = \sum_{k=1}^{t} \frac{\partial E_t}{\partial y_t} \frac{\partial y_t}{\partial h_t} \boxed{\frac{\partial h_t}{\partial h_k}} \frac{\partial h_k}{\partial W}$$

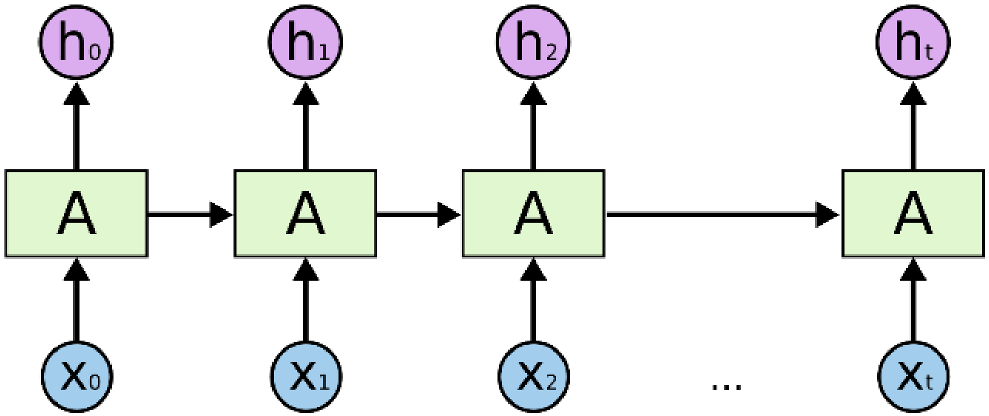
BP TT



$$\frac{\partial E_t}{\partial W} = \sum_{k=1}^{t} \frac{\partial E_t}{\partial y_t} \frac{\partial y_t}{\partial h_t} \boxed{\frac{\partial h_t}{\partial h_k}} \frac{\partial h_k}{\partial W}$$

$$\frac{\partial h_t}{\partial h_k} = \prod_{i=k+1}^{t} \frac{\partial h_i}{\partial h_{i-1}}$$

BP TT



$$\frac{\partial E_t}{\partial W} = \sum_{k=1}^{t} \frac{\partial E_t}{\partial y_t} \frac{\partial y_t}{\partial h_t} \boxed{\frac{\partial h_t}{\partial h_k}} \frac{\partial h_k}{\partial W}$$

$$\frac{\partial h_t}{\partial h_k} = \prod_{i=k+1}^{t} \frac{\partial h_i}{\partial h_{i-1}}$$

For example @ $t = 2$,

$$\frac{\partial h_2}{\partial h_0} = \prod_{i=1}^{2} \frac{\partial h_i}{\partial h_{i-1}} = \frac{\partial h_1}{\partial h_0} \frac{\partial h_2}{\partial h_1}$$

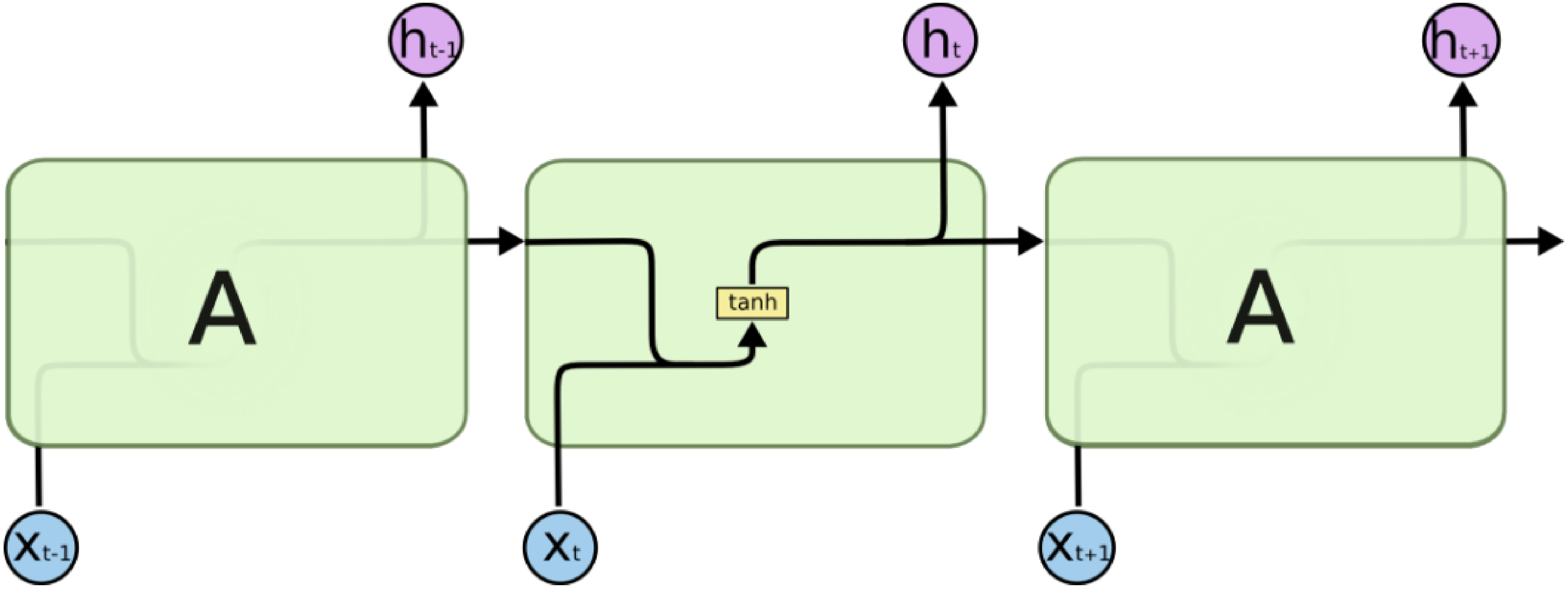
Vanishing (and Exploding) Gradients

- Exploding Gradients
  - Easy to detect
  - Clip the gradient at a threshold

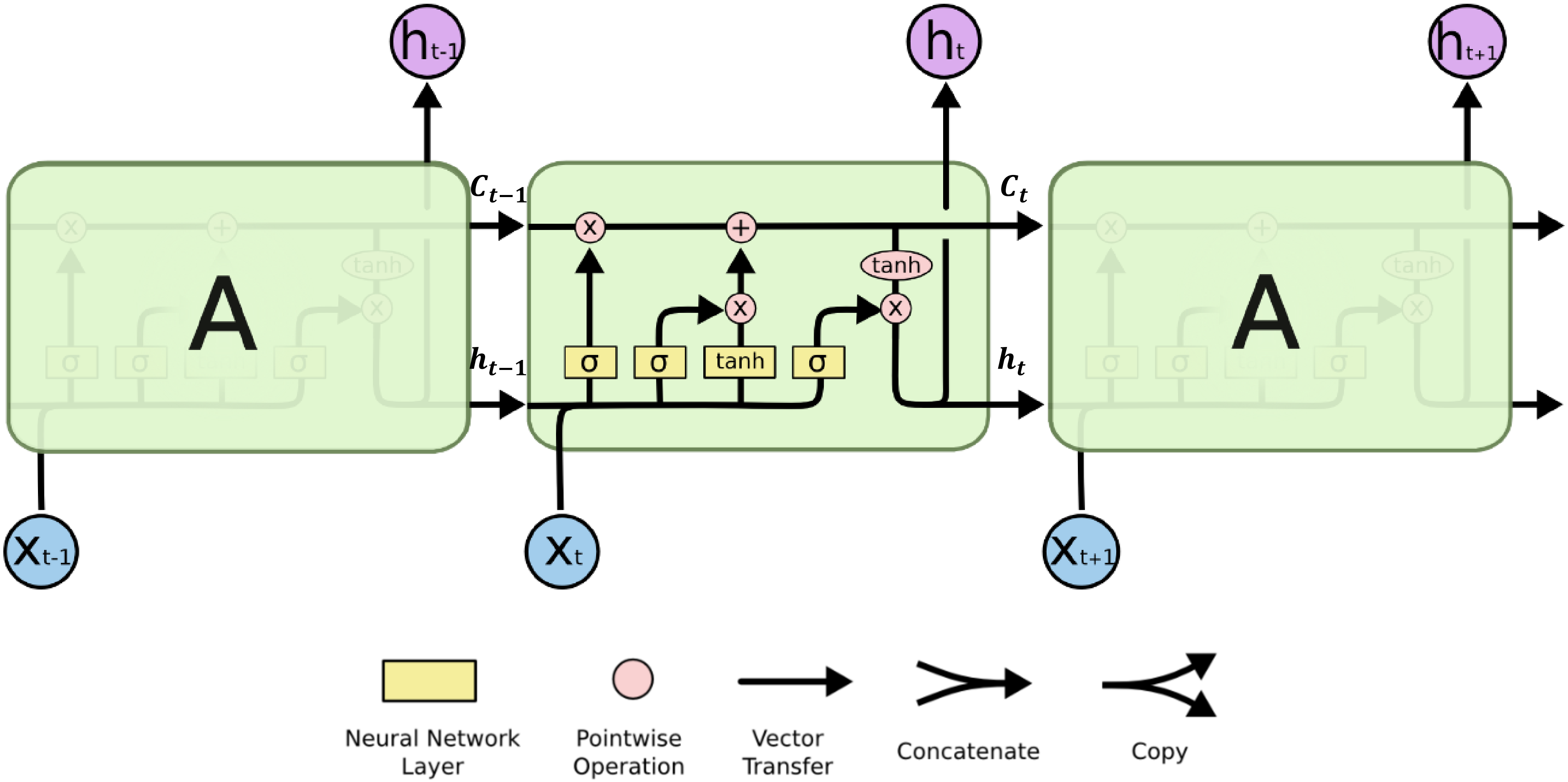
- Vanishing Gradients
  - More difficult to detect
  - Architectures designed to combat the problem of vanishing gradients. Example: LSTMs by *Schmidhuber et al.*

RNNs

- In a standard RNN the repeating module has a simple structure. Example:

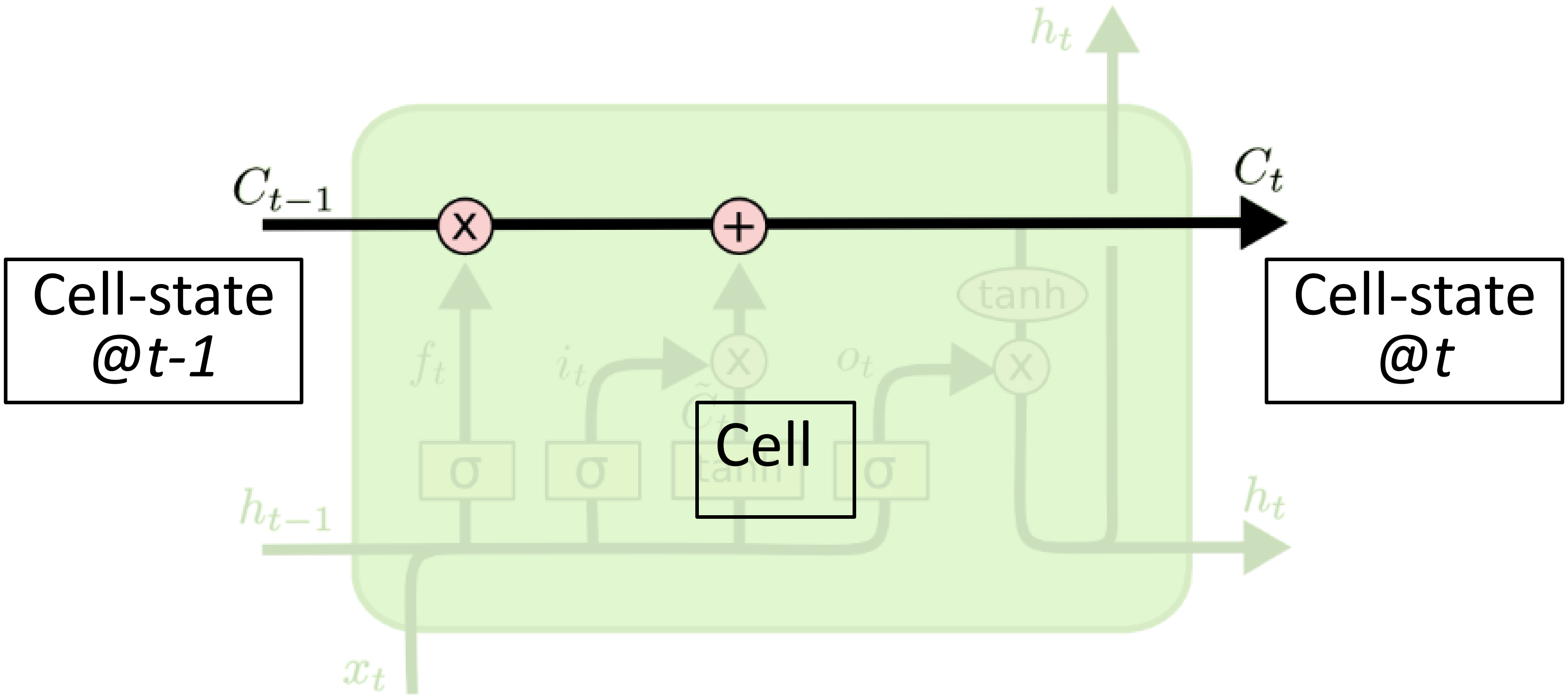


LSTMs



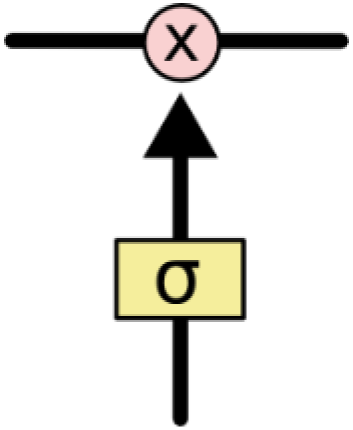
h t-1
h t
h t+1
A
A
C t−1
C t
h t−1
h t
σ
σ
tanh
σ
tanh
X t-1
X t
X t+1
Neural Network Layer
Pointwise Operation
Vector Transfer
Concatenate
Copy

LSTM Memory / Cell State



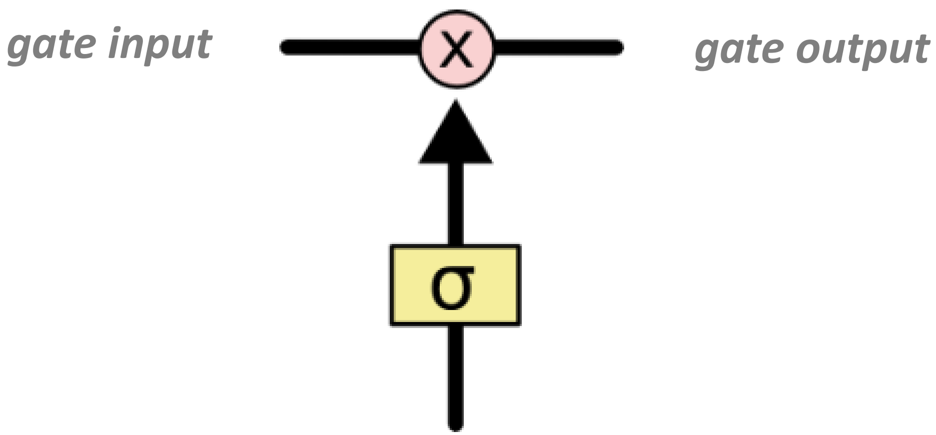
Gate

- Composed of a sigmoid neural net layer and a pointwise multiplication operation.



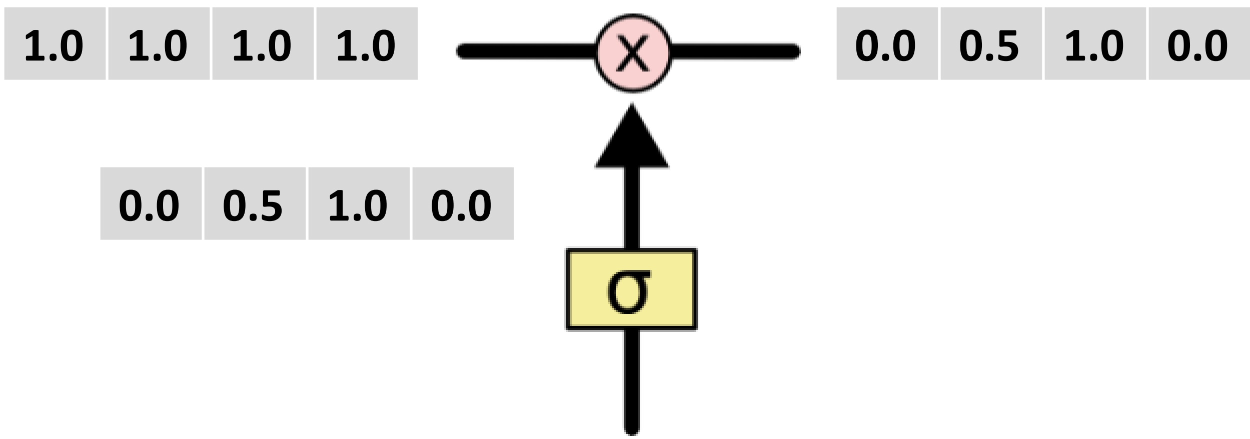
Gate

- sigmoid: outputs numbers between:
  - zero “let nothing through,” and
  - one, “let everything through!”
- Example:



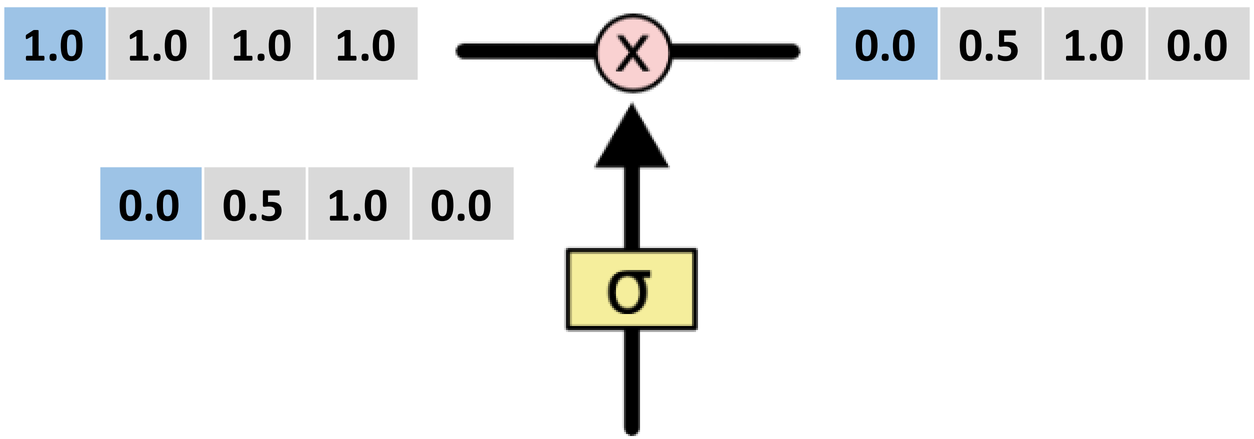
Gate

- sigmoid: outputs numbers between:
  - zero "let nothing through," and
  - one, "let everything through!"
- Example:



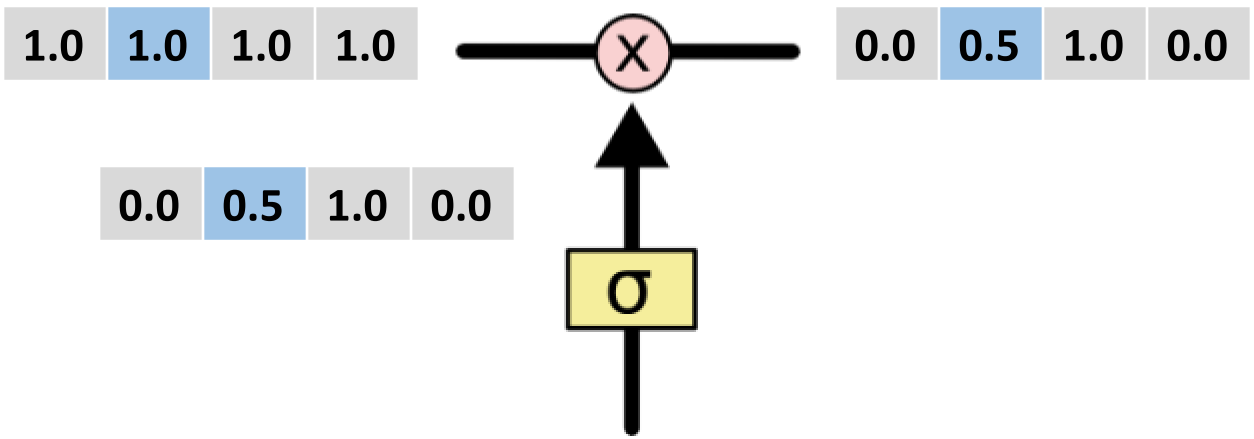
Gate

- sigmoid: outputs numbers between:
  - zero “let nothing through,” and
  - one, “let everything through!”
- Example:



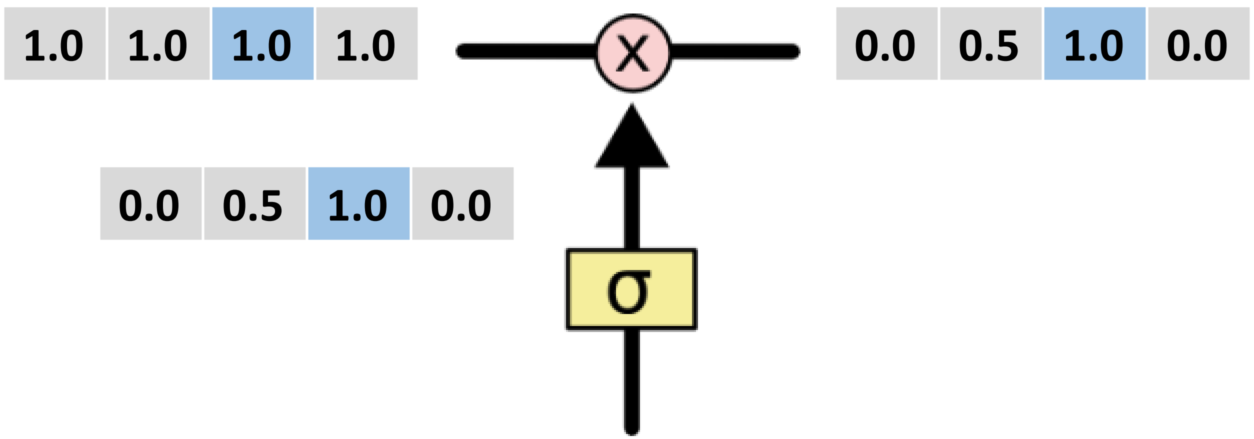
Gate

- sigmoid: outputs numbers between:
  - zero “let nothing through,” and
  - one, “let everything through!”
- Example:



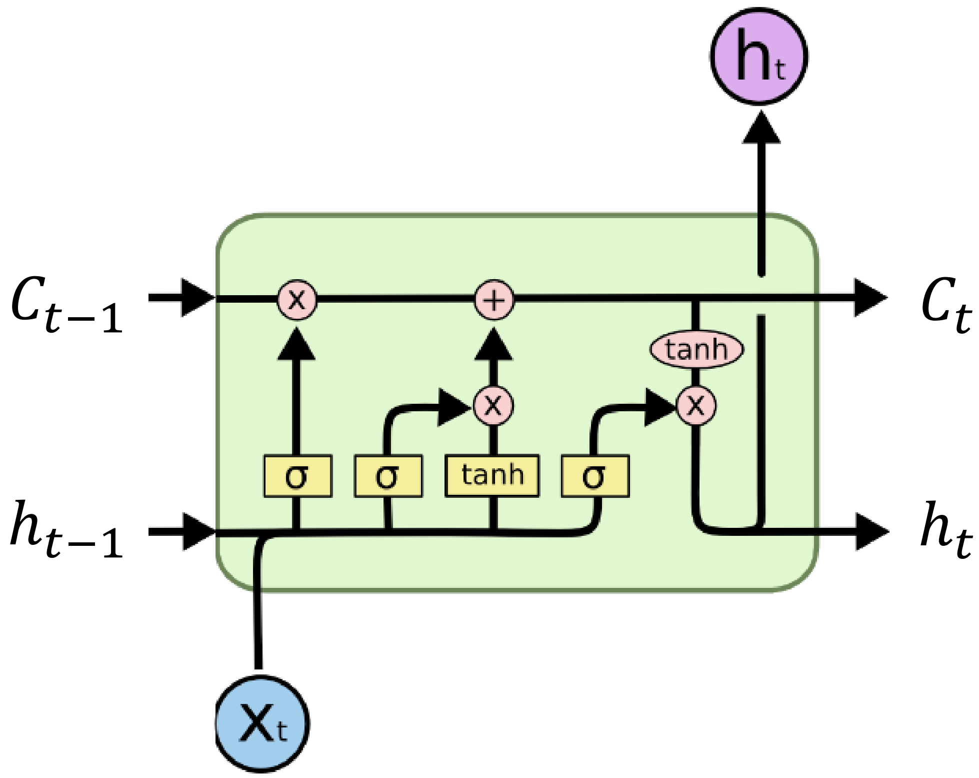
Gate

- sigmoid: outputs numbers between:
  - zero "let nothing through," and
  - one, "let everything through!"
- Example:



LSTM Gates

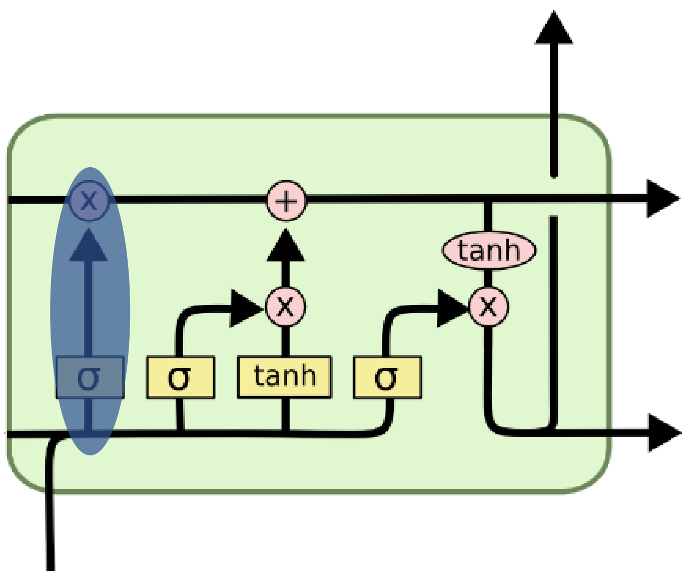
- An LSTM has three of these gates.



LSTM Gates

- An LSTM has three of these gates.

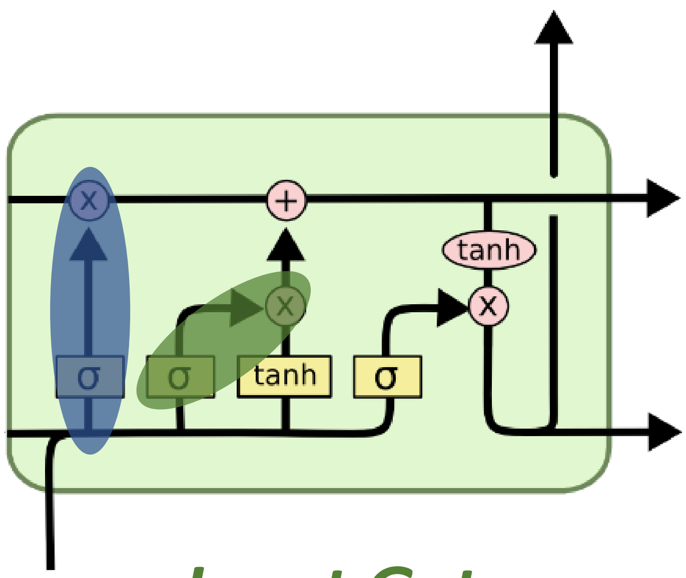
Forget gate



LSTM Gates

- An LSTM has three of these gates.

Forget gate

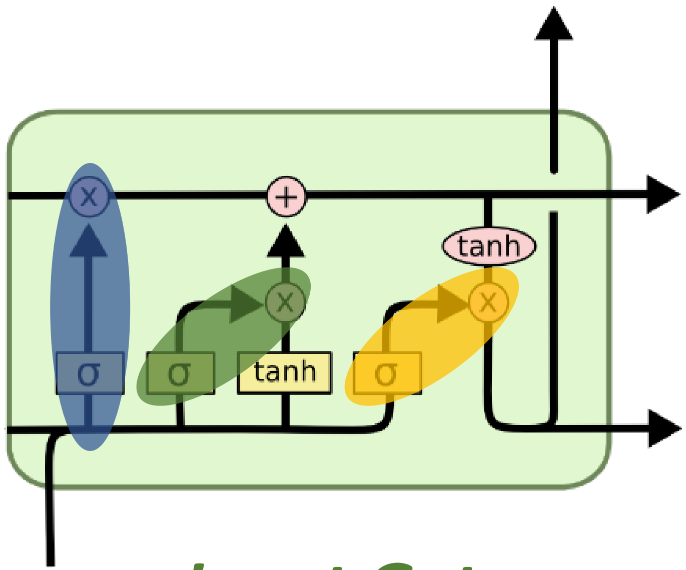


Input Gate

LSTM Gates

- An LSTM has three of these gates.

Forget gate



Output Gate

Input Gate

Example: AI Generated Trailer

- Analyze a movie and generate a trailer automatically

- How?
  Detecting salient moments
  e.g. action/emotions

- https://www.youtube.com/watch?v=gJEzuYynaiw

Detecting Salient Regions

- Two sample actions:

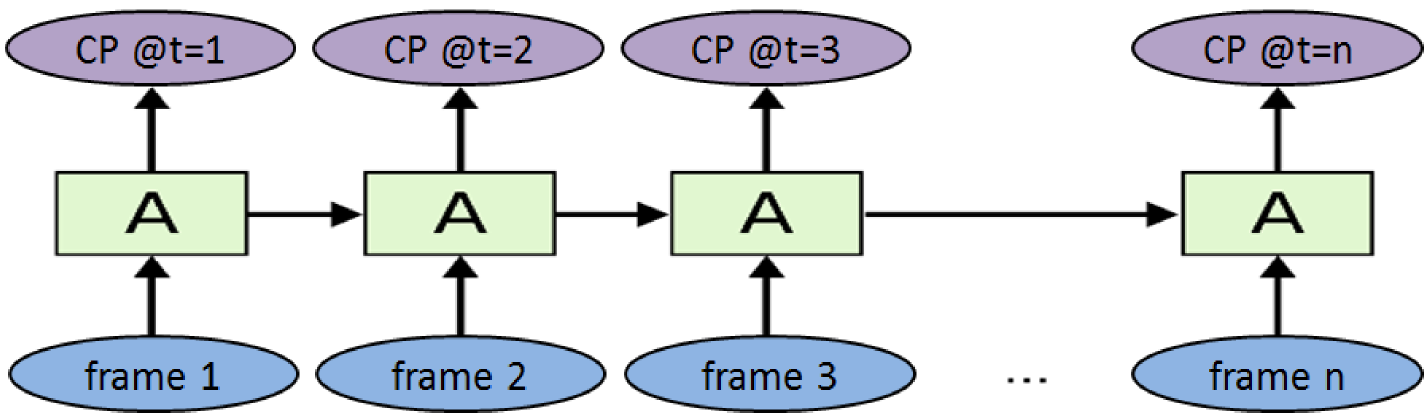
Handstand Walking

Ice Dancing

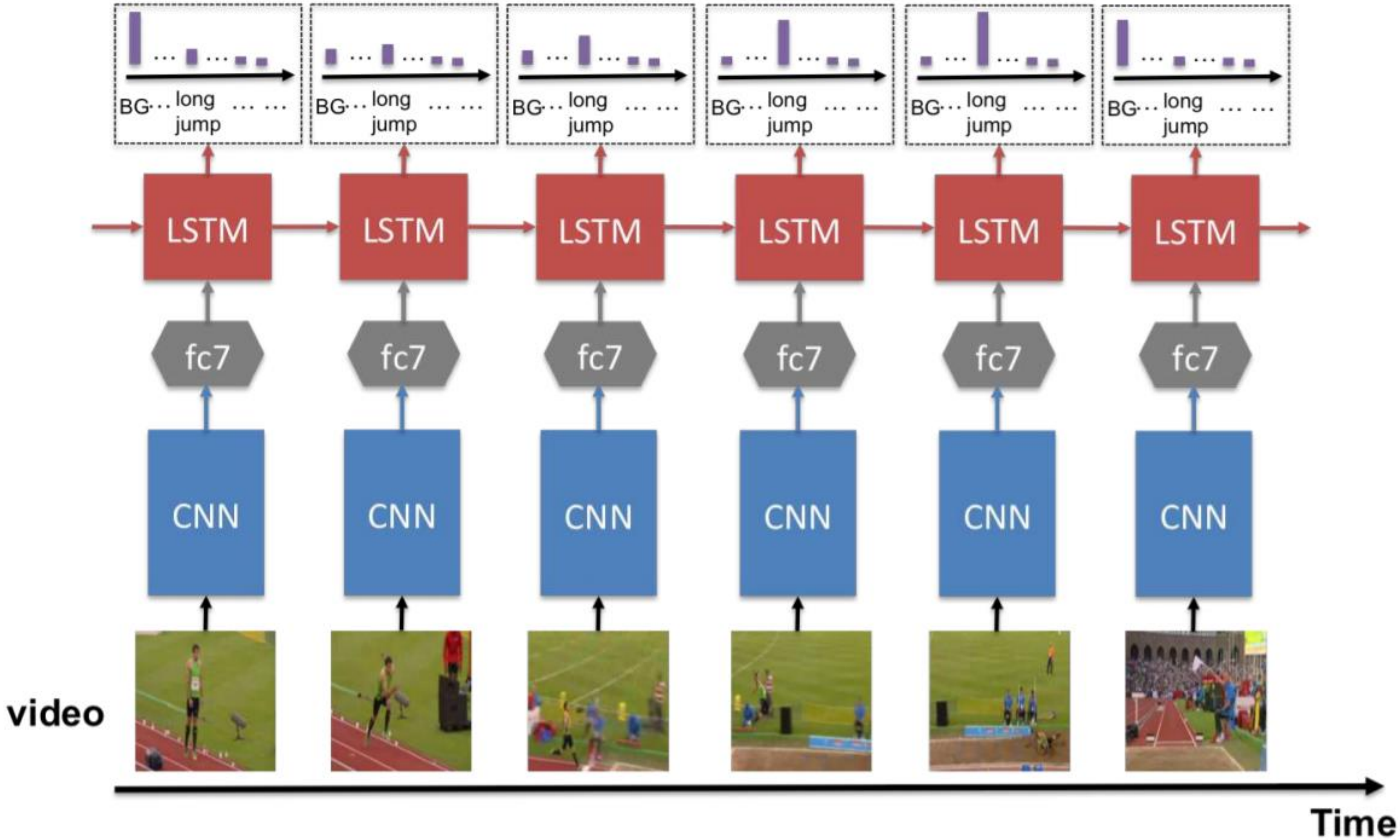
Applications of Recurrent Networks

Application 1: Video Classification

- CP: conditional class probability
- frame i could be a feature describing frame i, example: CNN feature



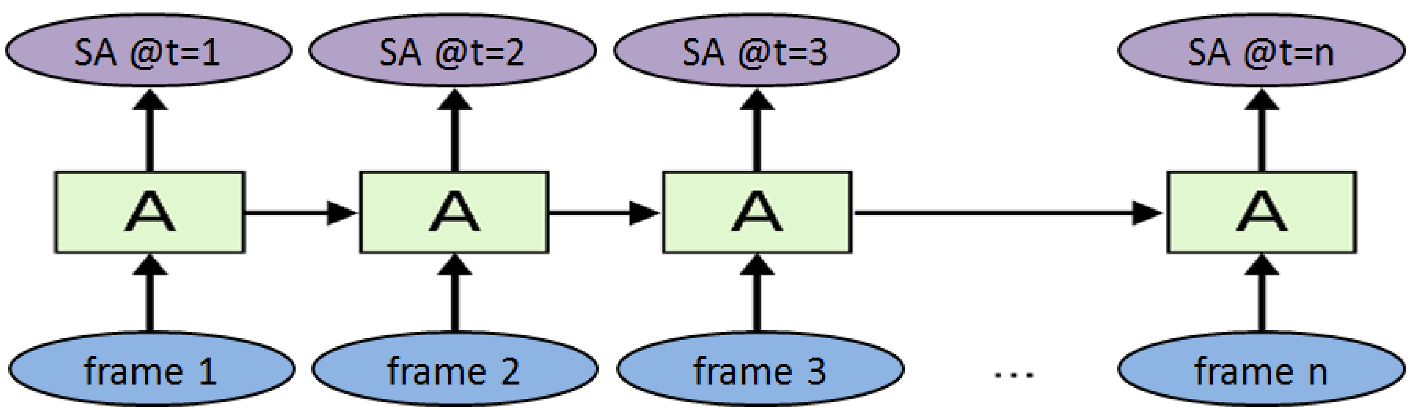
Application 1: Video Classification



Shugao Ma et al., CVPR'16, http://www.cv-foundation.org/openaccess/content_cvpr_2016/papers/Ma_Learning_Activity_Progression_CVPR_2016_paper.pdf

Application 2: Self-Driving Cars

- SA: steering angle
- frame i could be a feature describing frame i, example: 3D-CNN feature



Application 2: Self-Driving Cars

- DeepTesla

DeepTesla, http://selfdrivingcars.mit.edu/deeptesla/

Application 2: Self-Driving Cars

- Udacity winning team: *Team Komanda*
  - x_t: 3D convolution of image sequence
  - h_t: steering angle, speed, torque

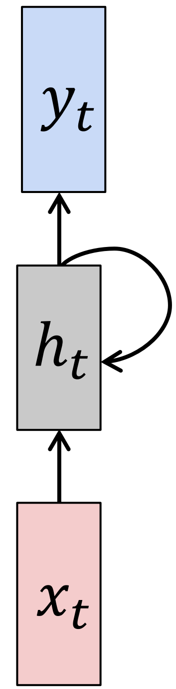
Lex Fridman, lecture, https://www.youtube.com/watch?v=nFTQ7kHQWtc

Application 3: Character RNN

Character-level language model example

Vocabulary:
[h,e,l,o]

Example training sequence:
“hello”

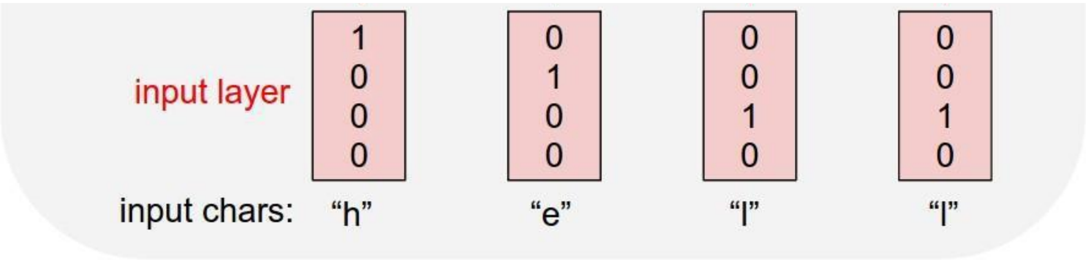


Application 3: Character RNN

Character-level language model example

Vocabulary: [h,e,l,o]

Example training sequence: **“hello”**

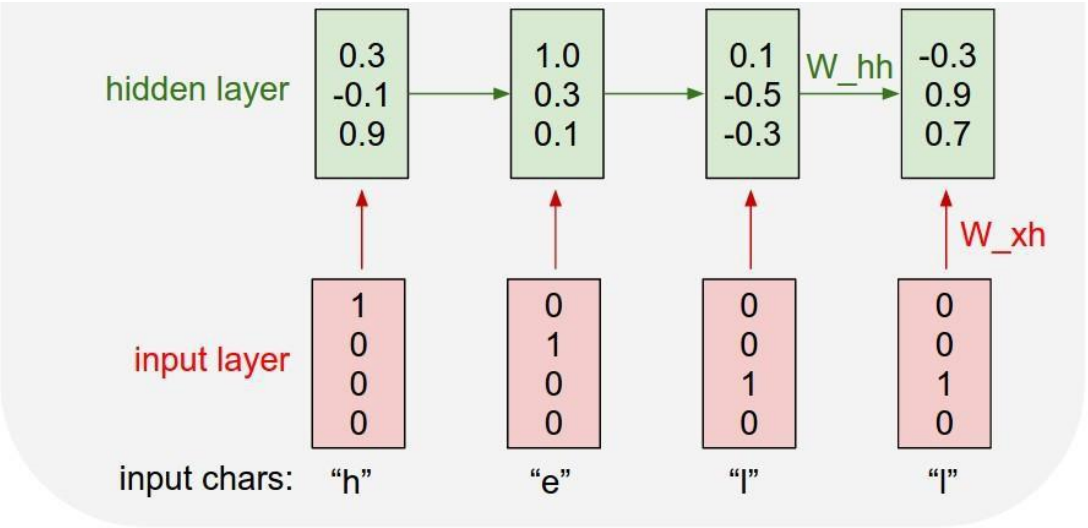


Application 3: Character RNN

Character-level language model example

Vocabulary: [h,e,l,o]

Example training sequence: **“hello”**

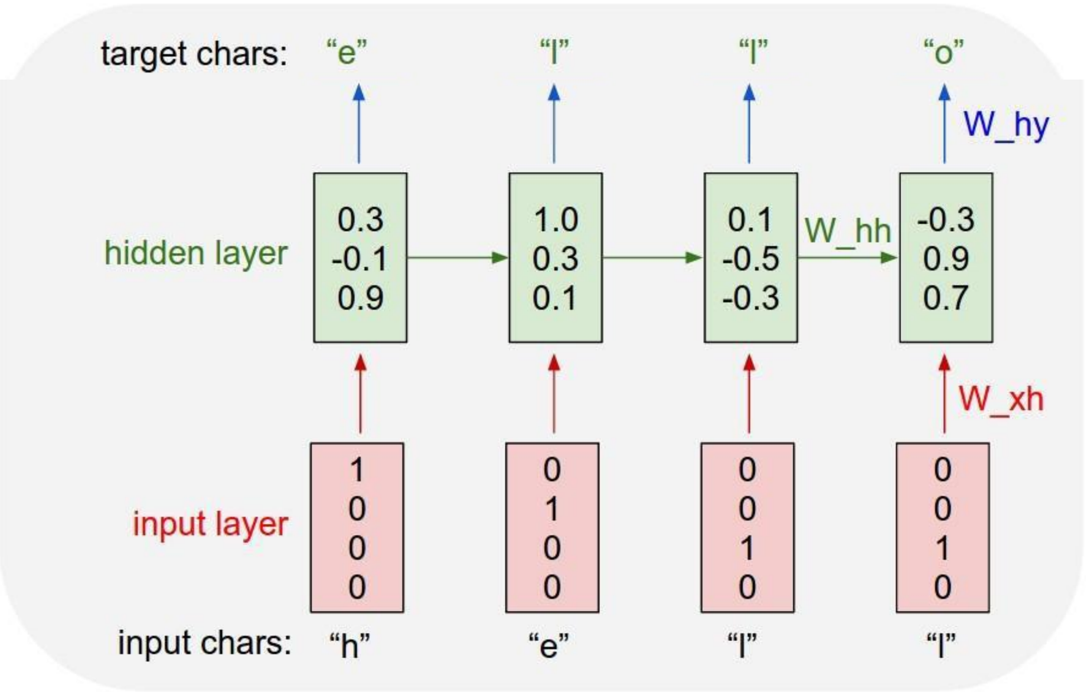


Application 3: Character RNN

Character-level language model example

Vocabulary:
[h,e,l,o]

Example training sequence:
“hello”



Application 4: Reading cursive

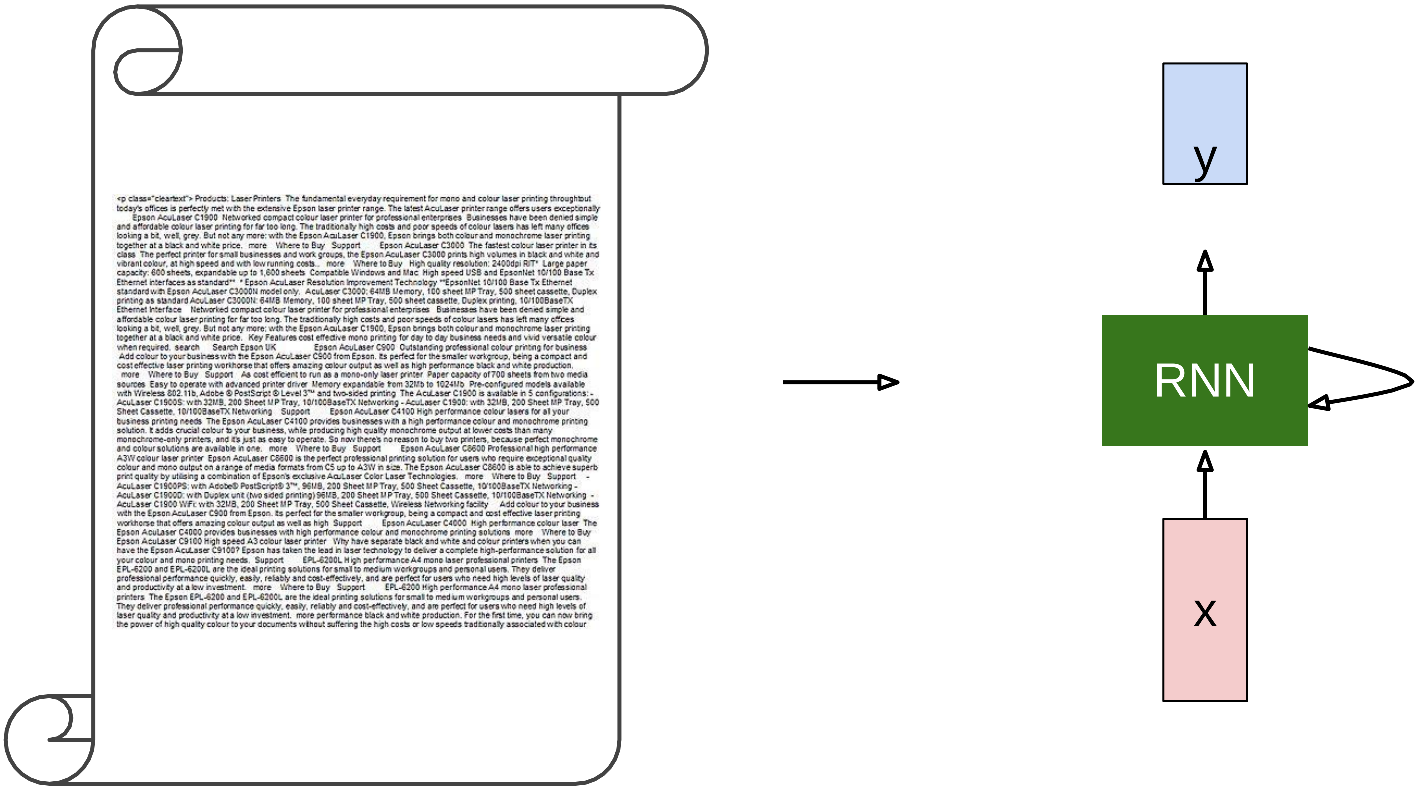
handwriting

- This is a natural task for an RNN.
- The input is a sequence of (x,y,p) coordinates of the tip of the pen, where p indicates whether the pen is up or down.
- The output is a sequence of characters.

- Graves & Schmidhuber (2009) showed that RNNs with LSTM are well-tailored for reading cursive writing.
    - They used a sequence of small images as input rather than pen coordinates.

Application 5: StyleText Generation

Training text: William Shakespeare



Application 5: StyleText Generation

at first:

```
tyntd-iafhatawiaoihrdemot  lytdws  e ,tfti, astai f ogoh eoase rrranbyne 'nhthnee e
plia tklrgd t o idoe ns,smtt   h ne etie h,hregtrs nigtike,aoaenns lng
```

train more

```
"Tmont thithey" fomesscerliund
Keushey. Thom here
sheulke, anmerenith ol sivh I lalterthend Bleipile shuwy fil on aseterlome
coaniogennc Phe lism thond hon at. MeiDimorotion in ther thize."
```

train more

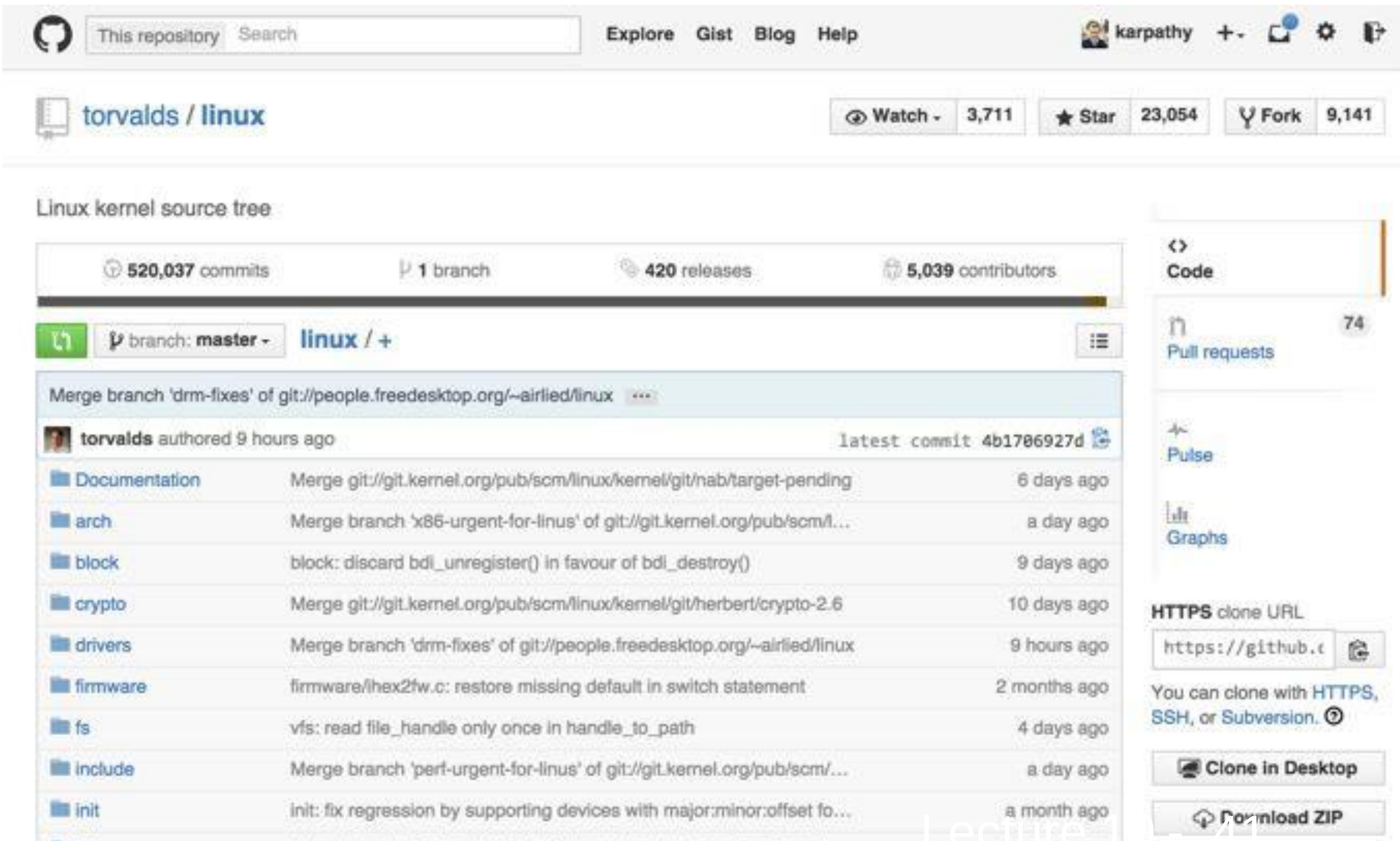
```
Aftair fall unsuch that the hall for Prince Velzonski's that me of
her hearly, and behs to so arwage fiving were to it beloge, pavu say falling misfort
how, and Gogition is so overelical and ofter.
```

train more

```
"Why do what that day," replied Natasha, and wishing to himself the fact the
princess, Princess Mary was easier, fed in had oftened him.
Pierre aking his soul came to the packs and drove up his father-in-law women.
```

Application 6: Code Generation

Train on C code



Application 6: Code Generation

```c
static void do_command(struct seq_file *m, void *v)
{
  int column = 32 << (cmd[2] & 0x80);
  if (state)
    cmd = (int)(int_state ^ (in_8(&ch->ch_flags) & Cmd) ? 2 : 1);
  else
    seq = 1;
  for (i = 0; i < 16; i++) {
    if (k & (1 << 1))
      pipe = (in_use & UMXTHREAD_UNCCA) +
        ((count & 0x00000000fffffff8) & 0x000000f) << 8;
    if (count == 0)
      sub(pid, ppc_md.kexec_handle, 0x20000000);
    pipe_set_bytes(i, 0);
  }
  /* Free our user pages pointer to place camera if all dash */
  subsystem_info = &of_changes[PAGE_SIZE];
  rek_controls(offset, idx, &soffset);
  /* Now we want to deliberately put it to device */
  control_check_polarity(&context, val, 0);
  for (i = 0; i < COUNTER; i++)
    seq_puts(s, "policy ");
}
```

Generated C code

Application 7: Writing a Movie Script

https://arstechnica.com/the-multiverse/2016/06/an-ai-wrote-this-movie-and-its-strangely-moving/