

Sequence Data and Recurrent Neural Networks

Neural Networks Design And Application

Convolutional neural networks

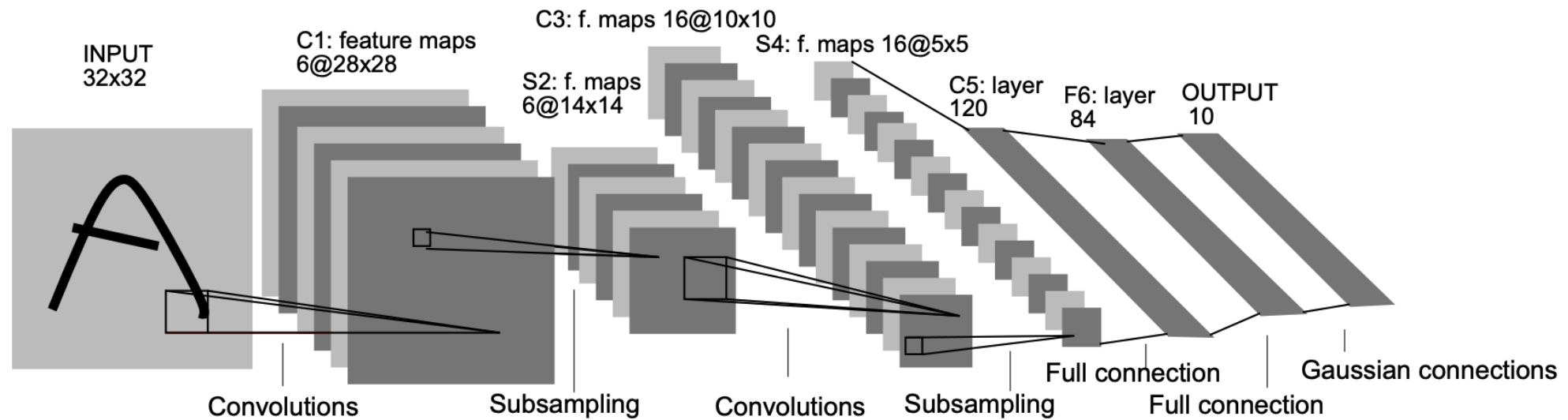


Fig. 1. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

Convolutional neural networks

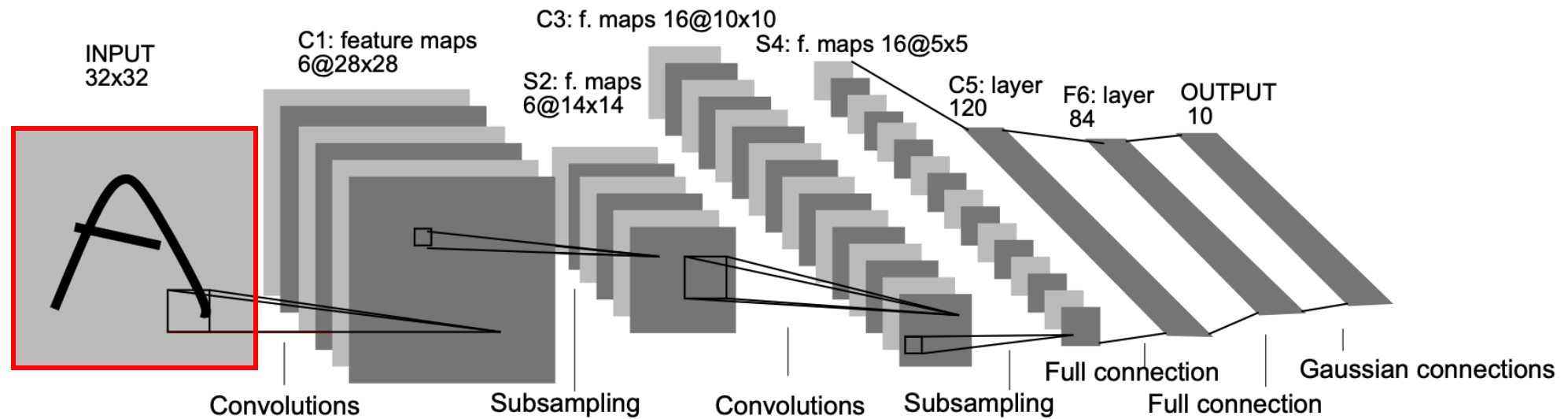


Fig. 1. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

Convolutional neural networks



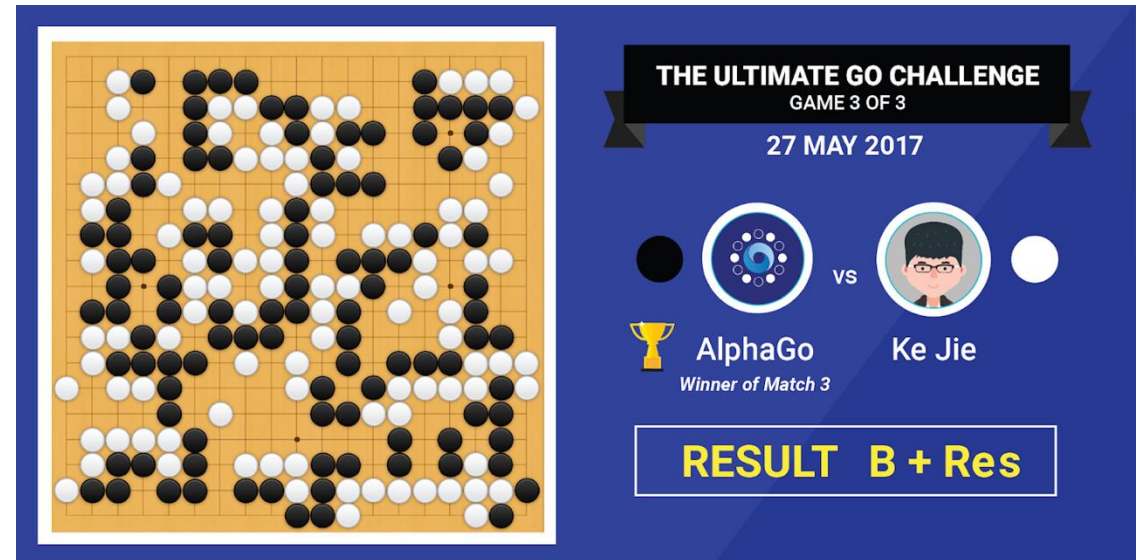
Fig. 1. Architecture of a convolutional neural network for digit recognition. Each plane in the network is constrained to be identical.

Convolutional neural networks

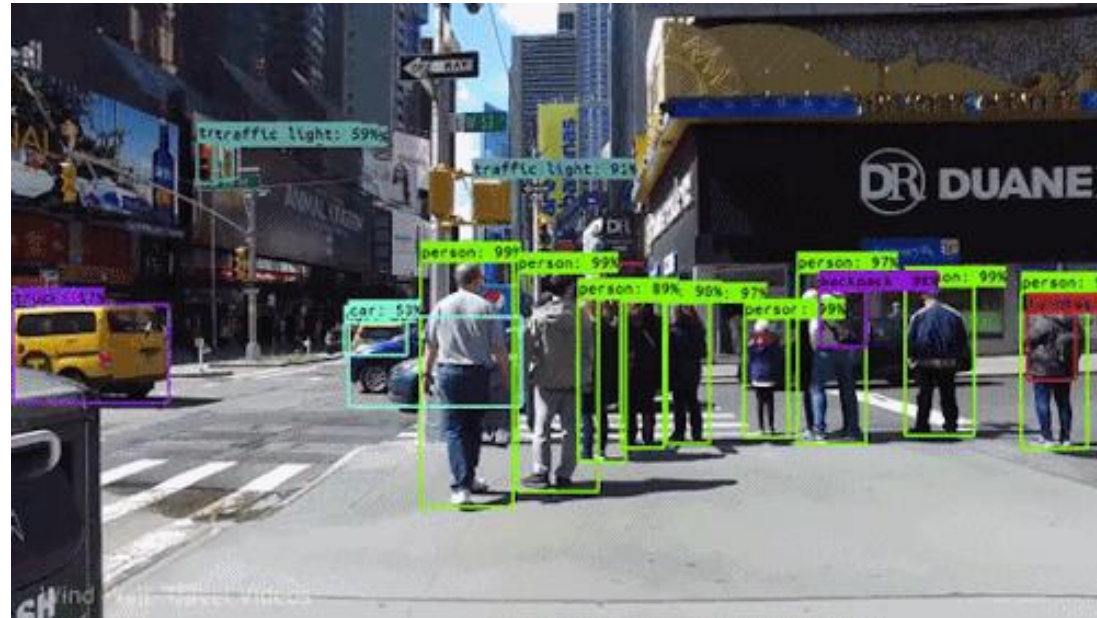


Fig. 1. Architecture of a convolutional neural network for digit recognition. Each plane in the network is required to be identical.

Some data may not be independent



Some data may not be independent



Some data may not be independent

[A demo video](https://pjreddie.com/darknet/yolo/) of **YOLOv3** from <https://pjreddie.com/darknet/yolo/>

Limitations of FC nets and CNNs

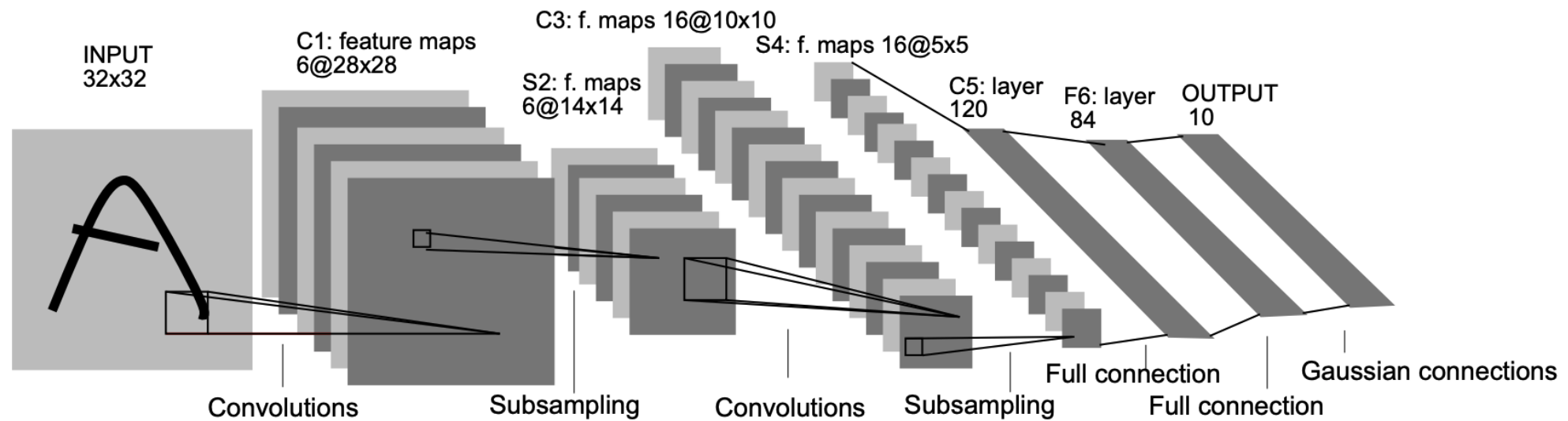


Fig. 1. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

Limitations of FC nets and CNNs

one to one

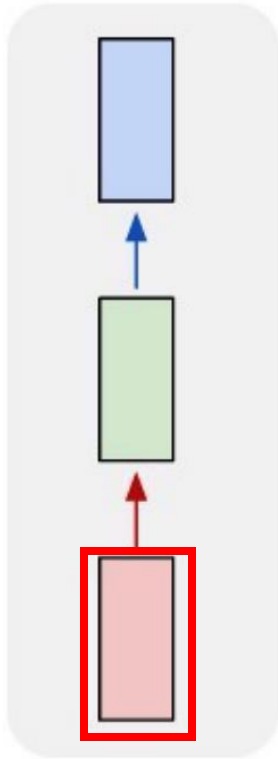


Image data: a single sample

Limitations of FC nets and CNNs

one to one

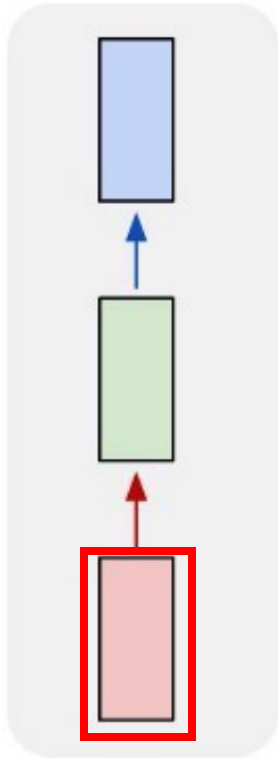


Image data: a single sample

Q: what if video data (e.g., 60 frame per second)?

Limitations of FC nets and CNNs

one to one

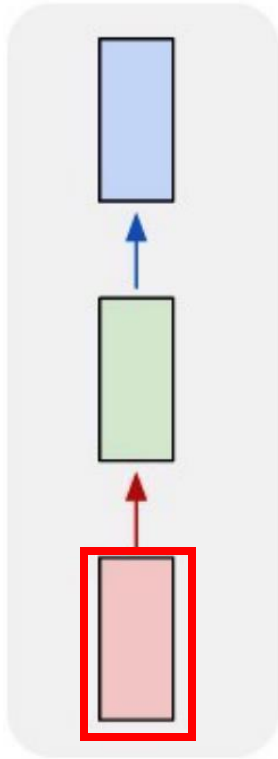
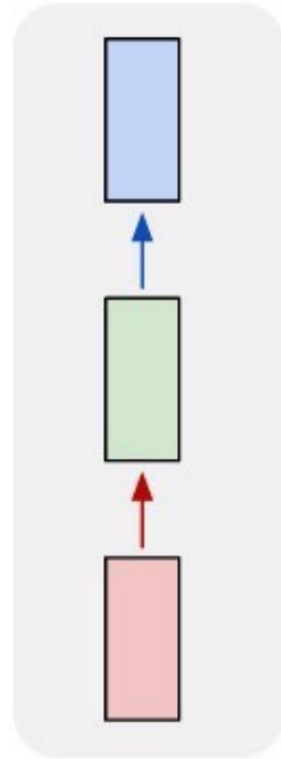


Image data: a single sample

Q: what if video data (e.g., 60 frame per second)?

one to one



Video data: multiple frames per second

Limitations of FC nets and CNNs

one to one

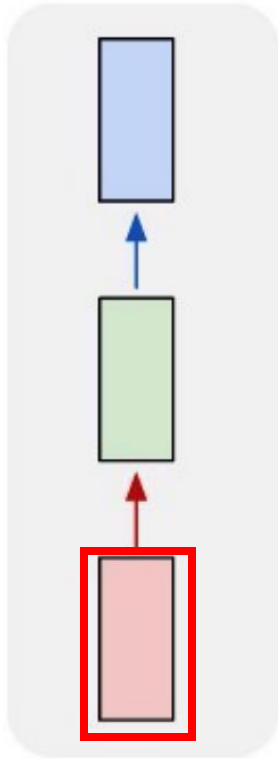
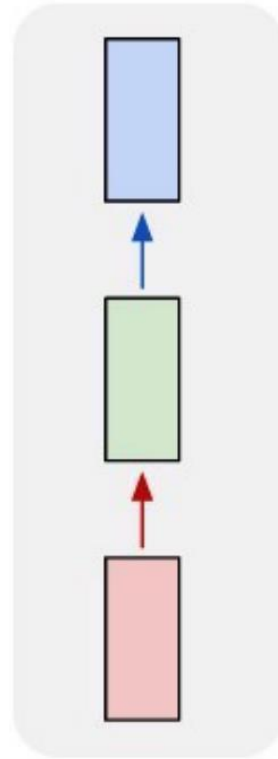


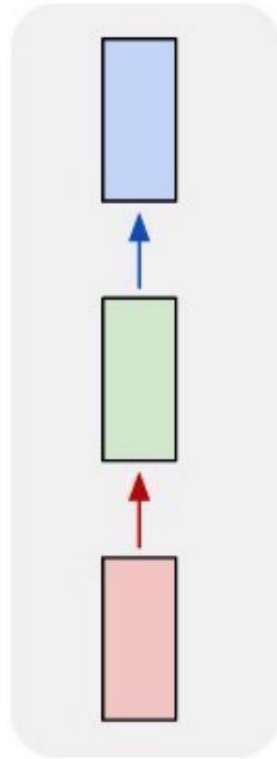
Image data: a single sample

Q: what if video data (e.g., 60 frame per second)?

one to one



one to one



Video data: multiple frames per second

Limitations of FC nets and CNNs

one to one

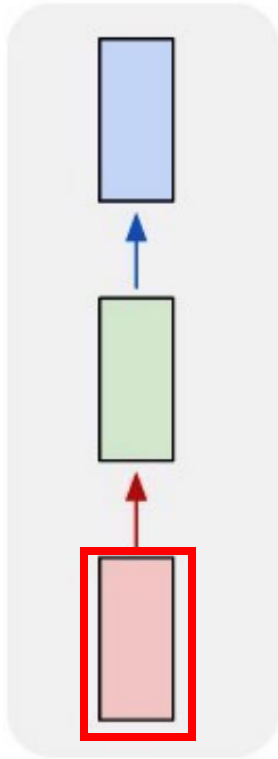
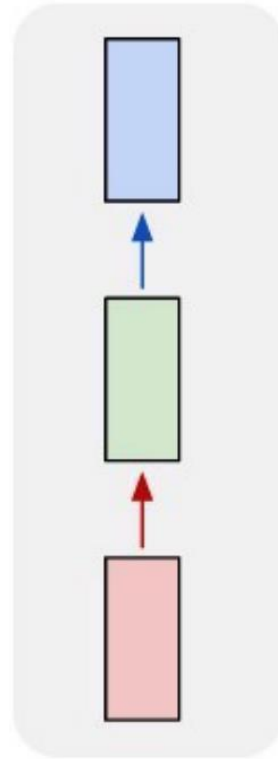


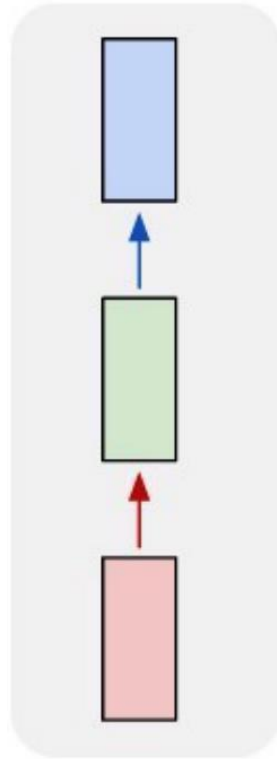
Image data: a single sample

Q: what if video data (e.g., 60 frame per second)?

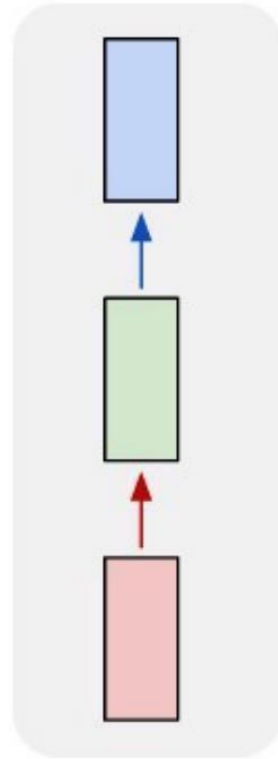
one to one



one to one



one to one



Video data: multiple frames per second

Limitations of FC nets and CNNs

one to one

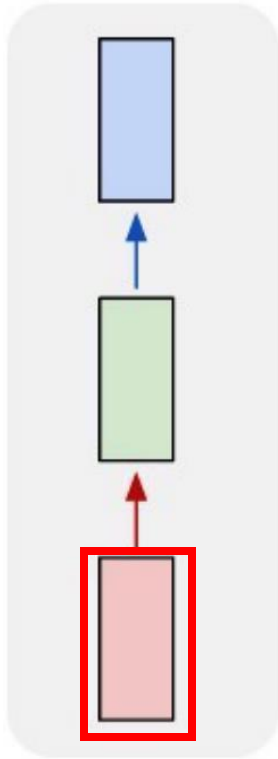
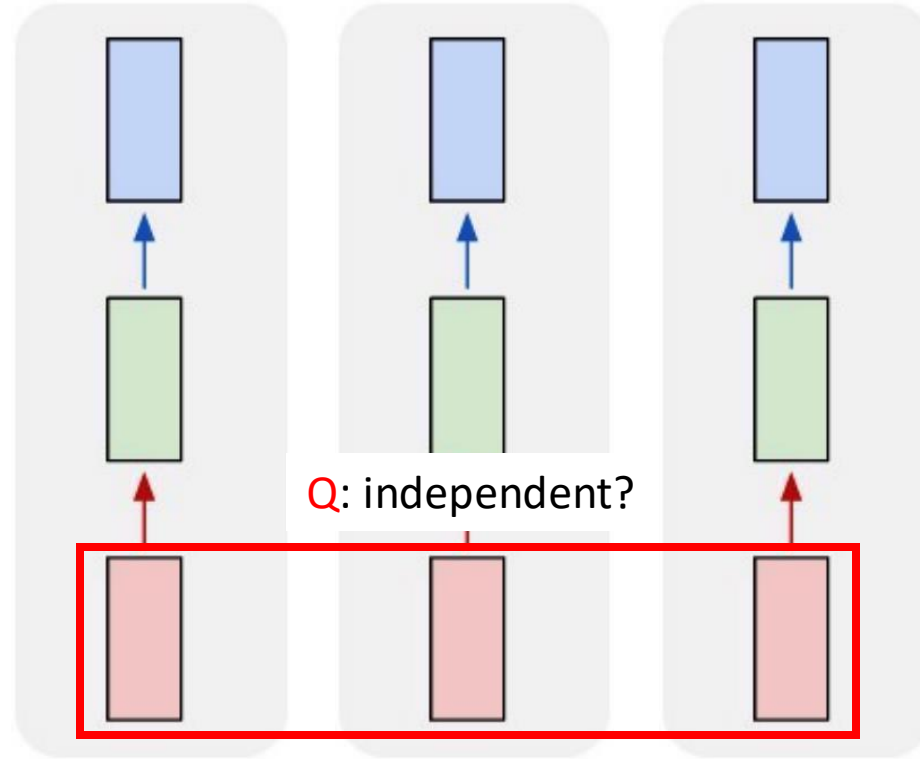


Image data: a single sample

Q: what if video data (e.g., 60 frame per second)?

one to one one to one one to one



Video data: multiple frames per second

Limitations of FC nets and CNNs

one to one

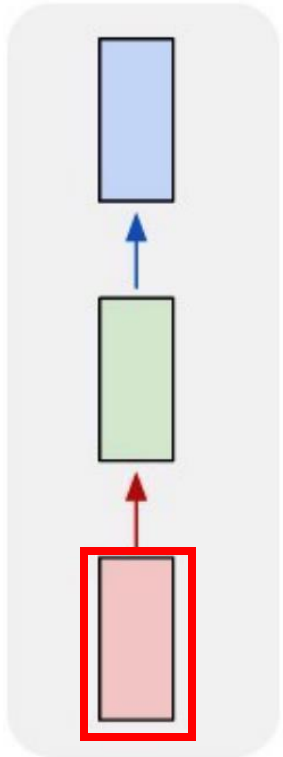
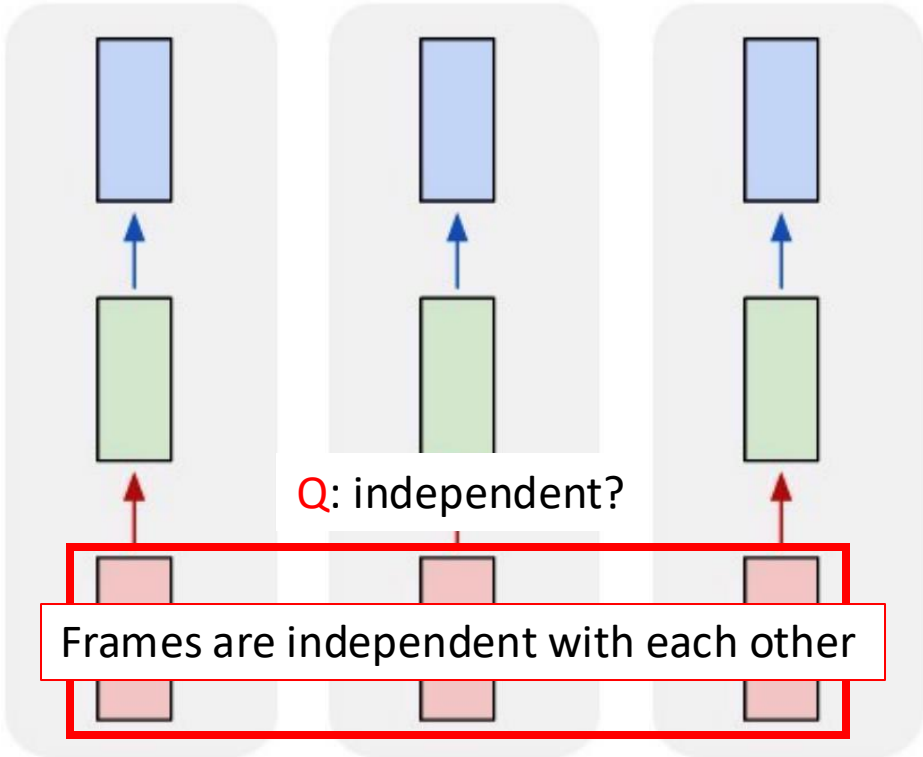


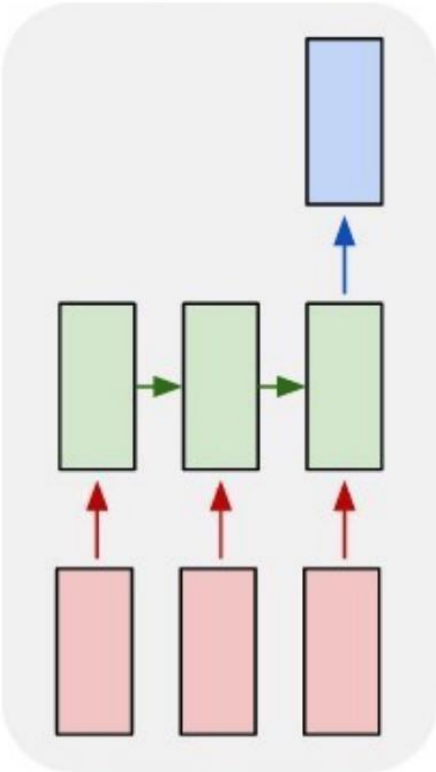
Image data: a single sample
Q: what if video data (e.g., 60 frame per second)?

one to one one to one one to one



Video data: multiple frames per second

many to one



Video data: multiple frames per second

Limitations of FC nets and CNNs

one to one

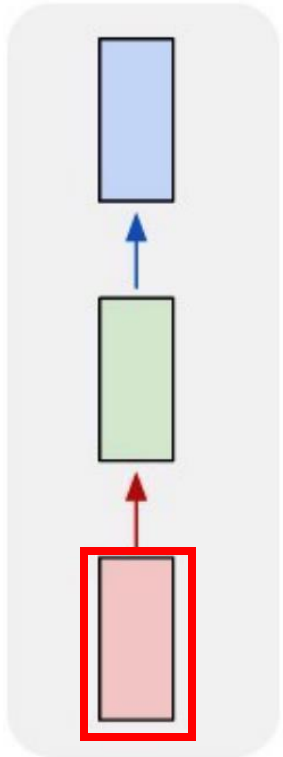
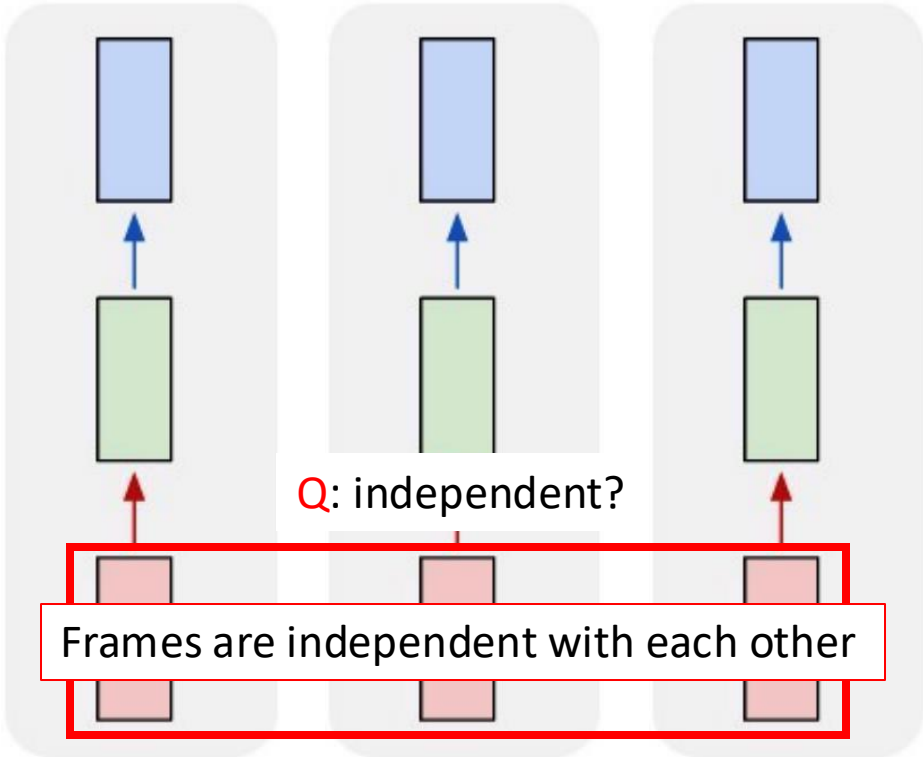


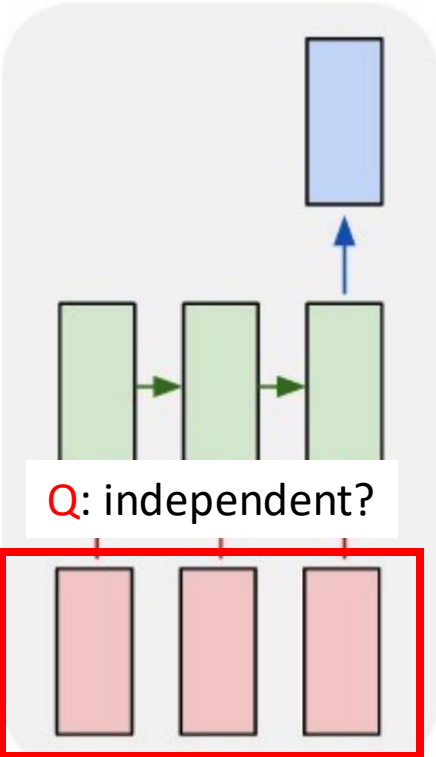
Image data: a single sample
Q: what if video data (e.g., 60 frame per second)?

one to one one to one one to one



Video data: multiple frames per second

many to one



Video data: multiple frames per second

Limitations of FC nets and CNNs

one to one

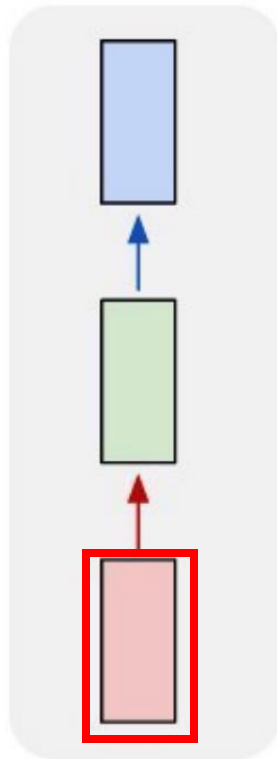
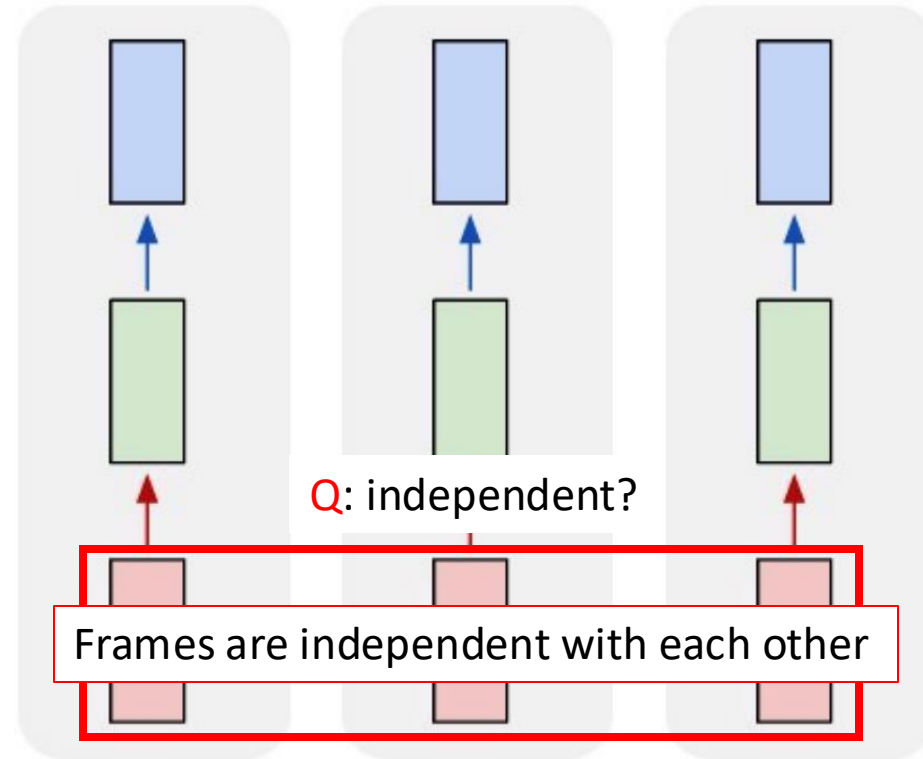


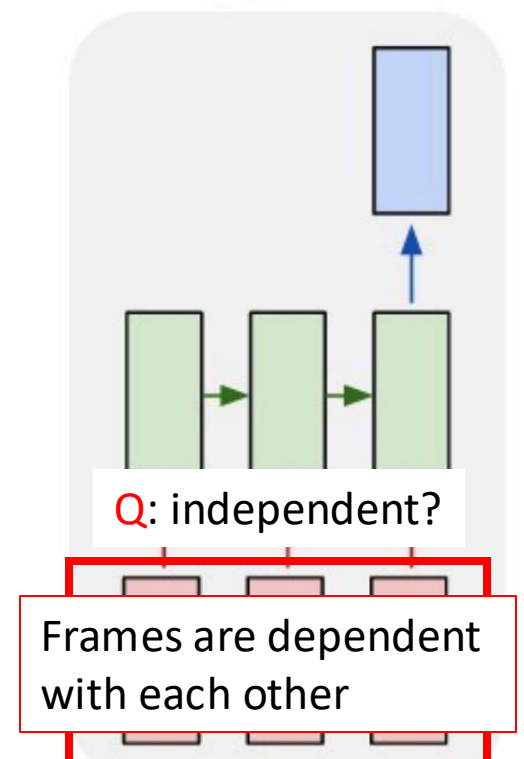
Image data: a single sample
Q: what if video data (e.g., 60 frame per second)?

one to one one to one one to one



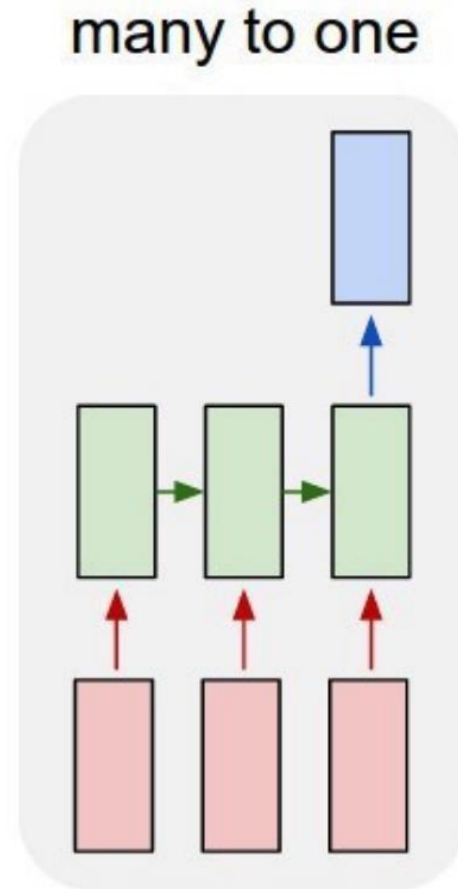
Video data: multiple frames per second

many to one

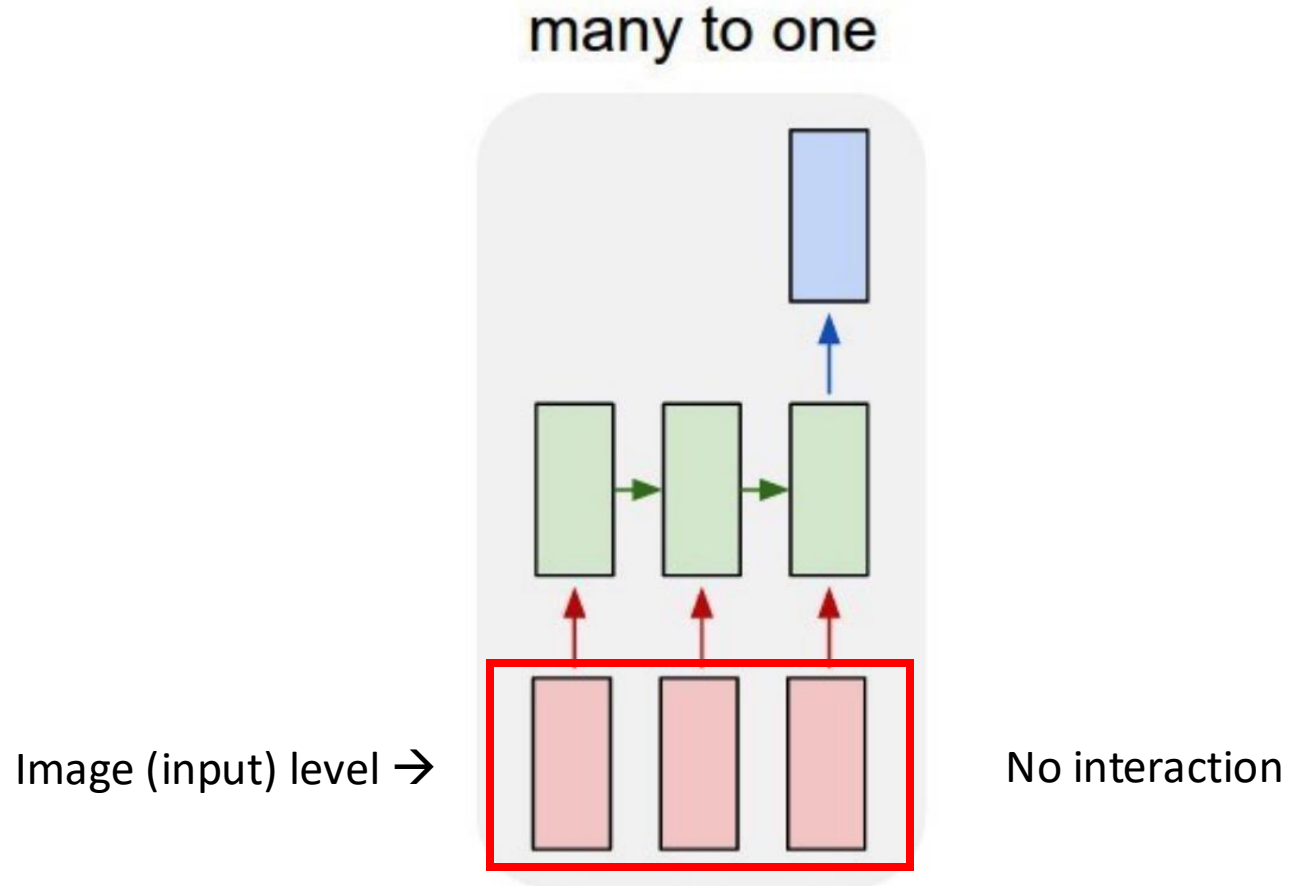


Video data: multiple frames per second

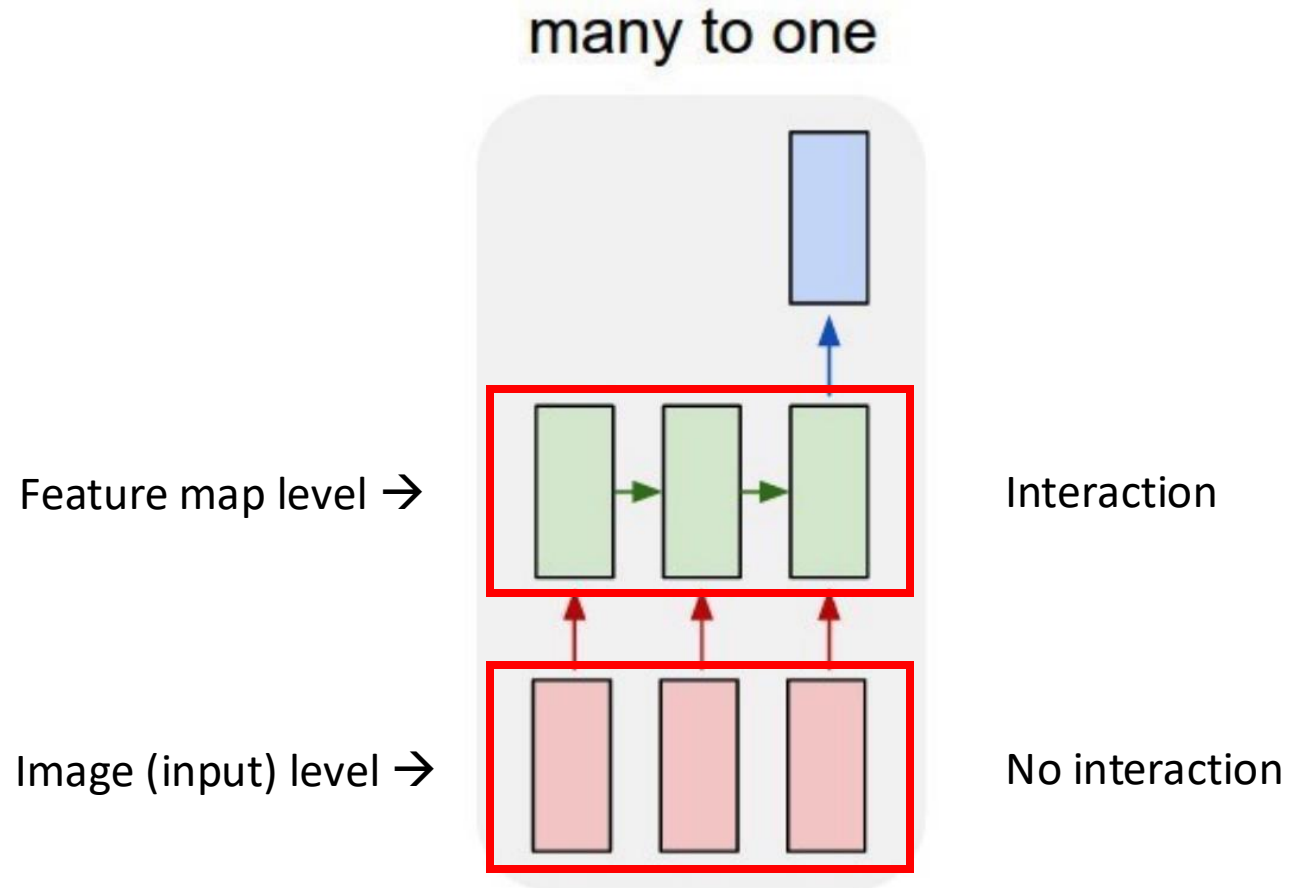
Limitations of FC nets and CNNs



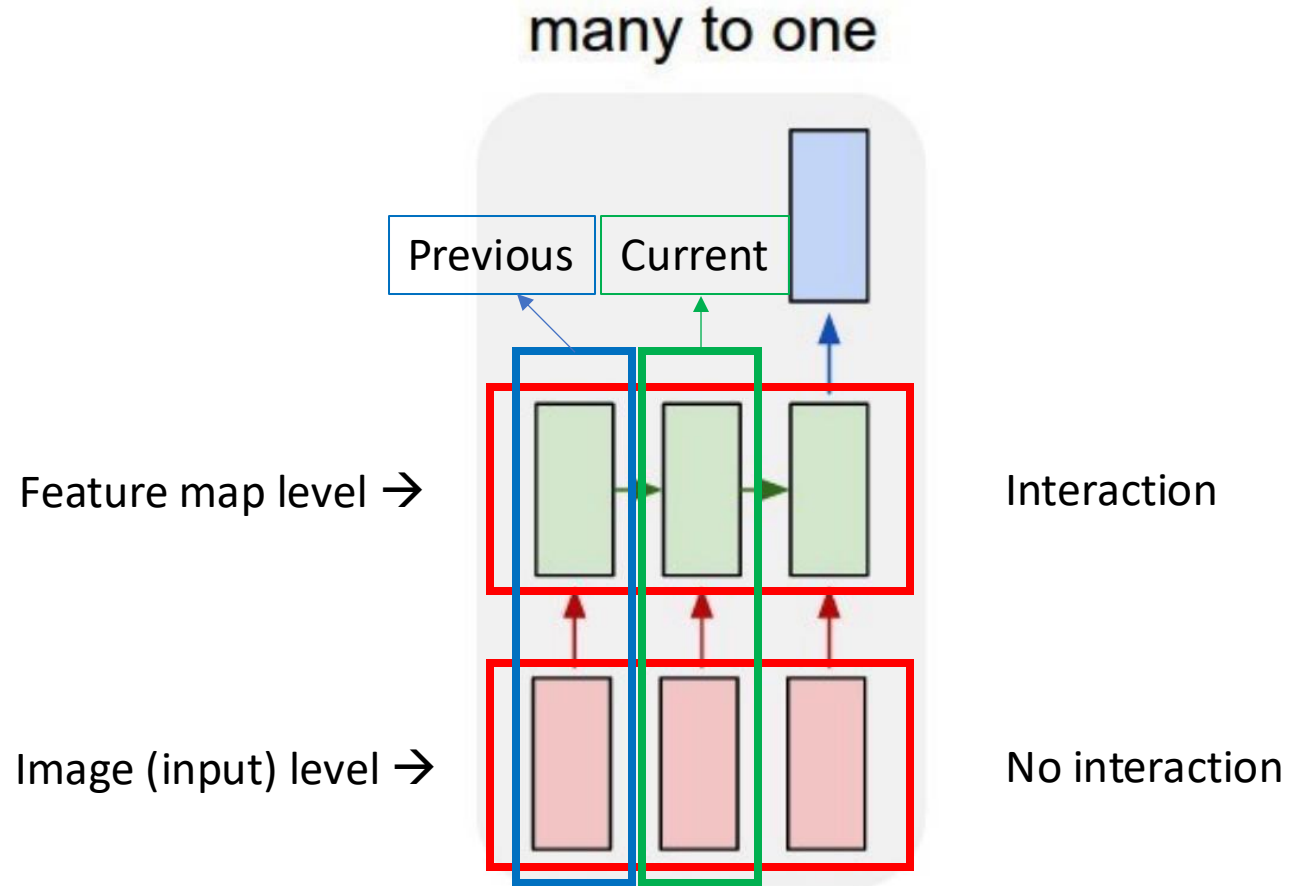
Limitations of FC nets and CNNs



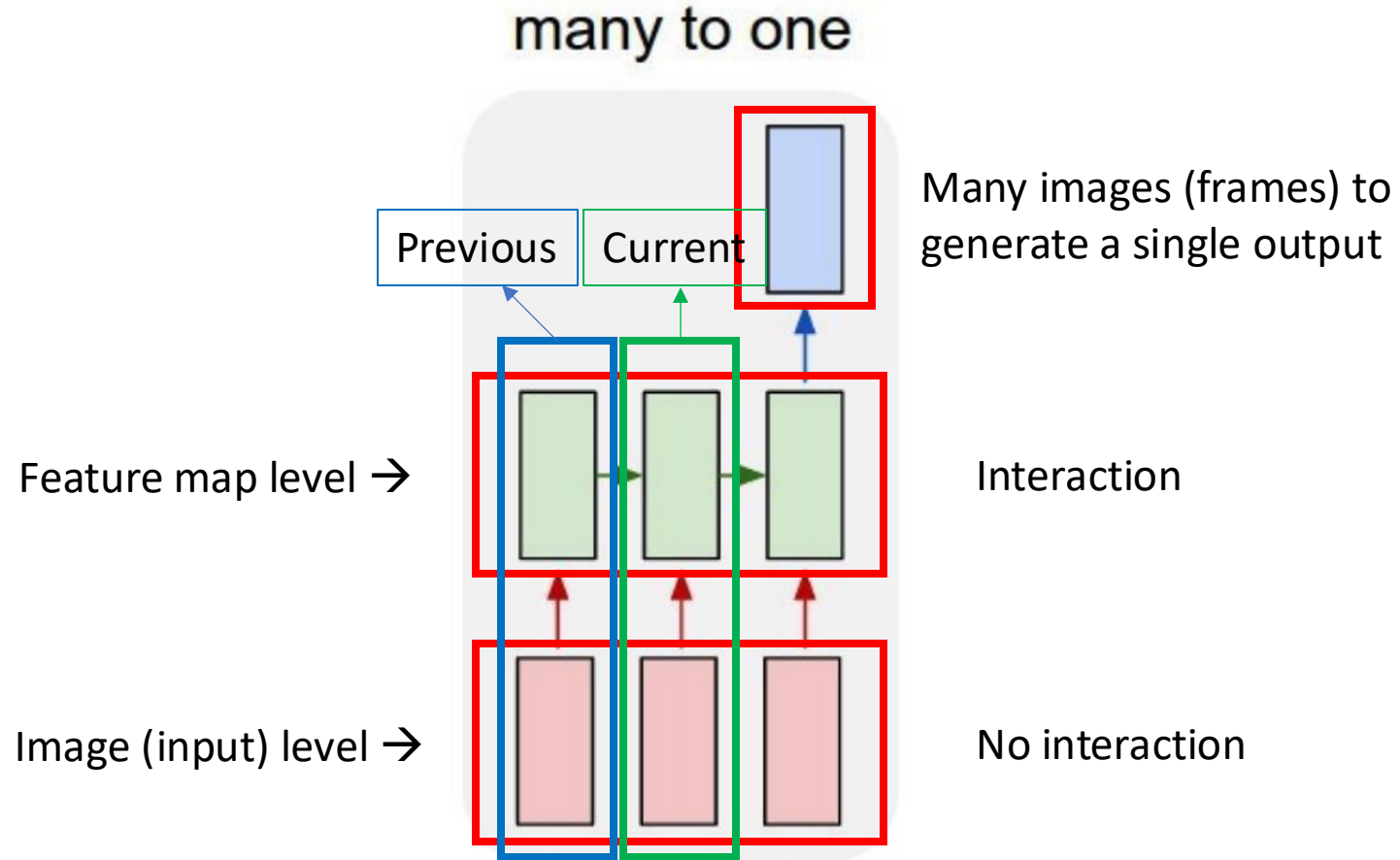
Limitations of FC nets and CNNs



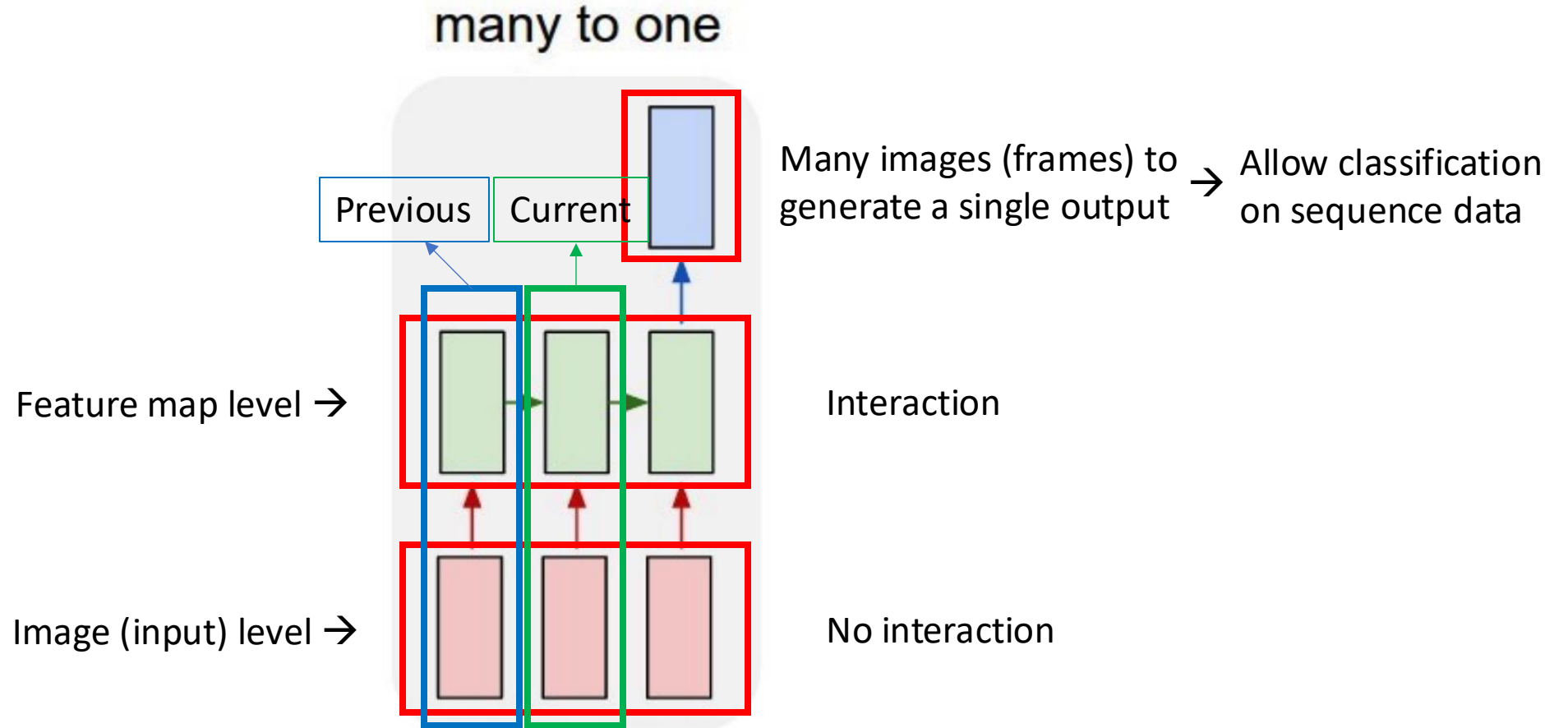
Limitations of FC nets and CNNs



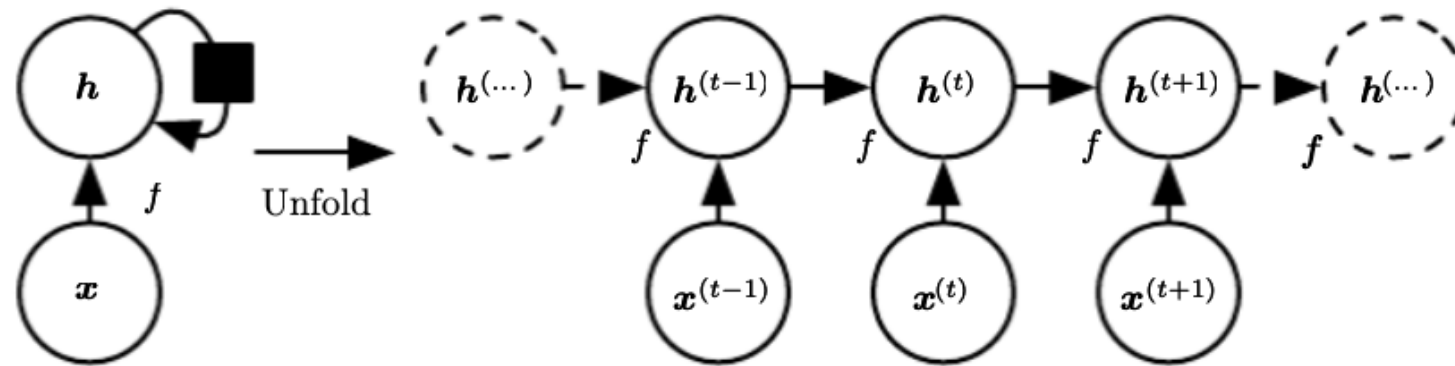
Limitations of FC nets and CNNs



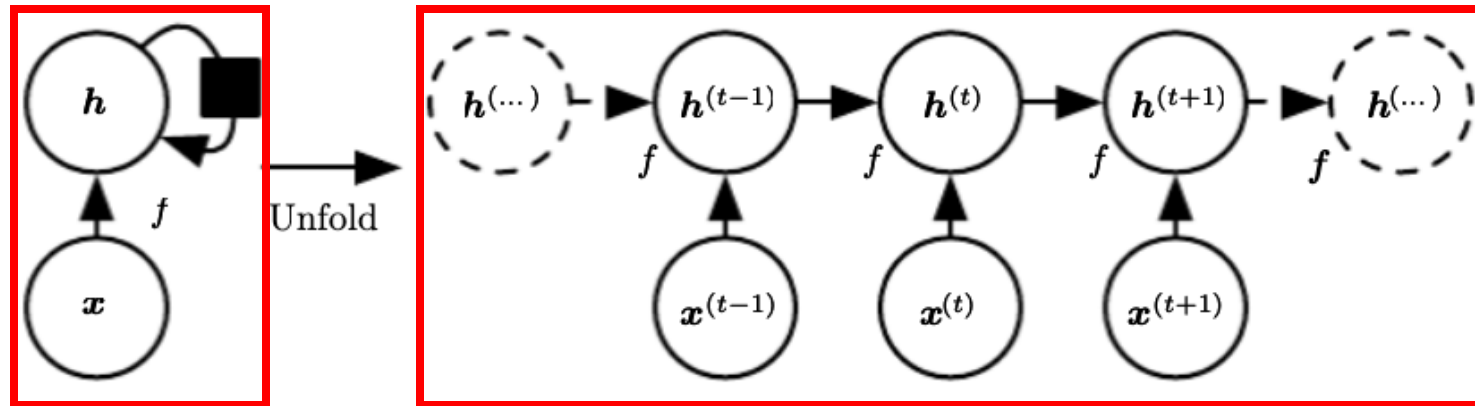
Limitations of FC nets and CNNs



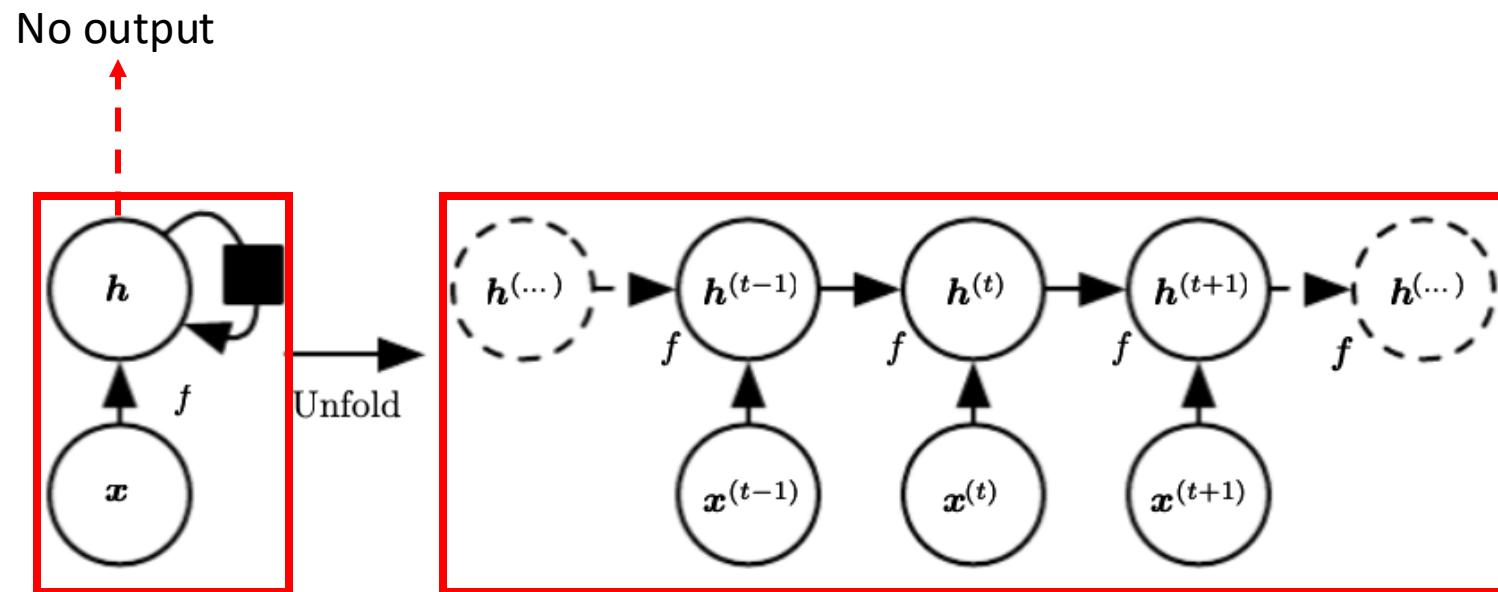
Recurrent networks



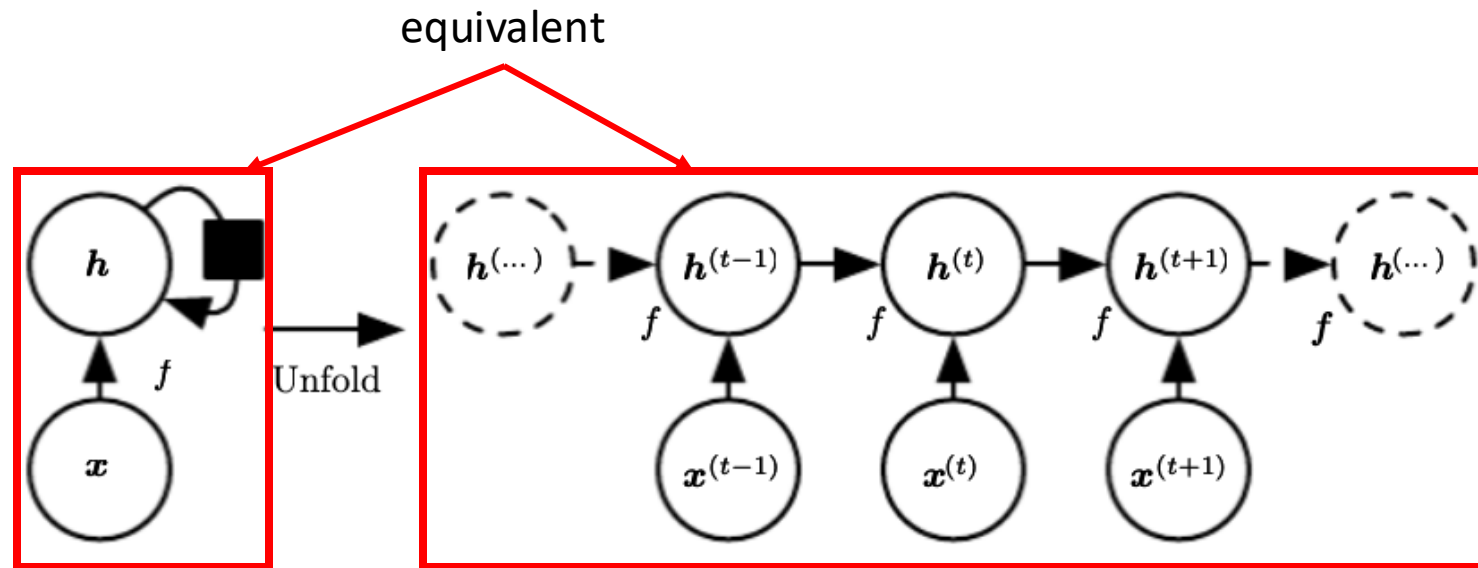
Recurrent networks



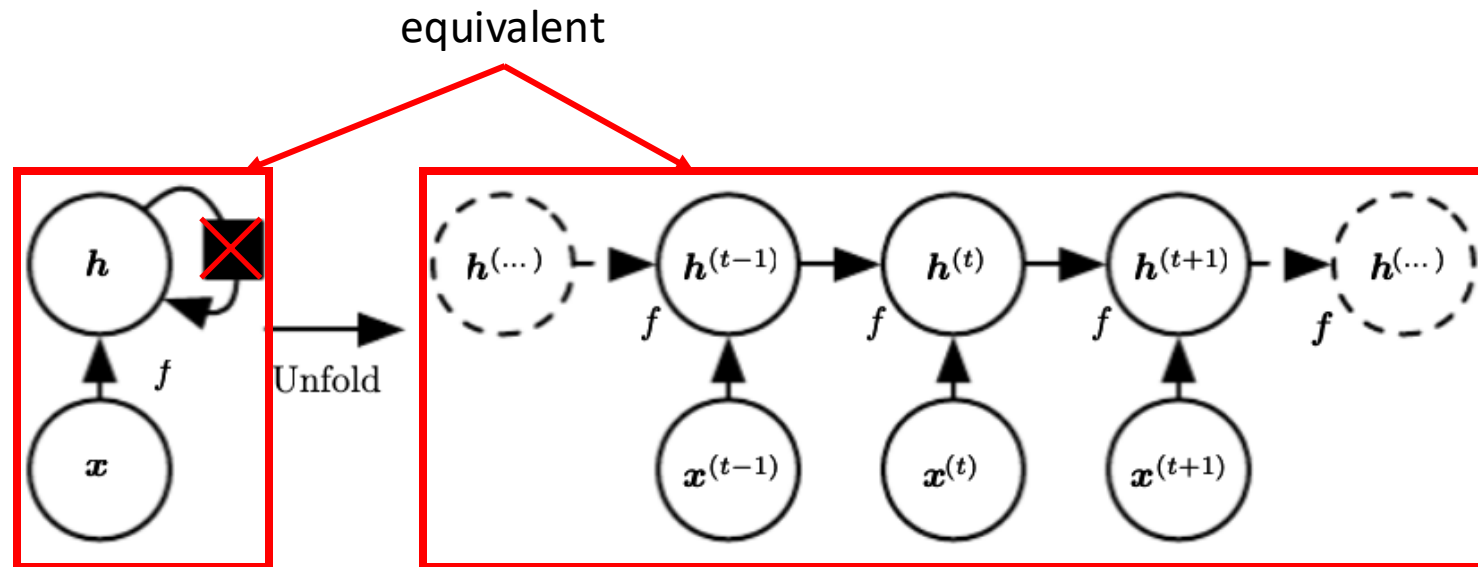
Recurrent networks



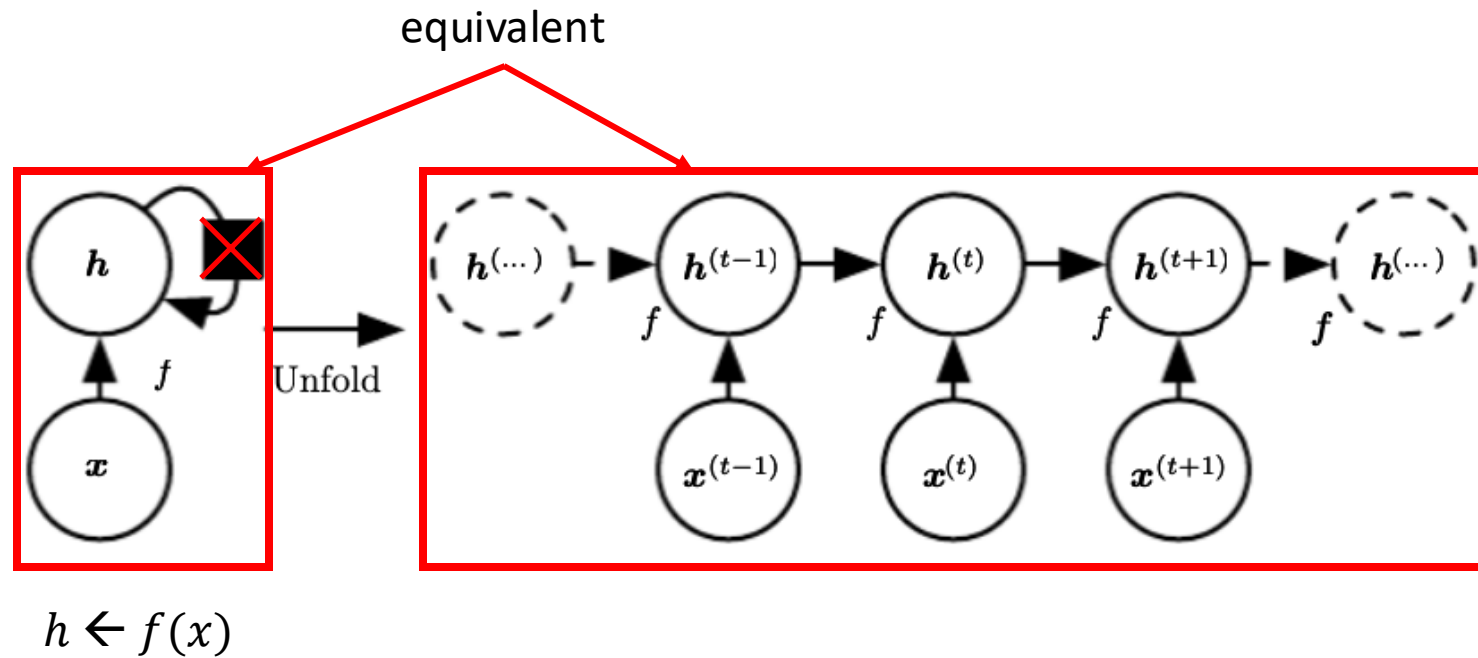
Recurrent networks



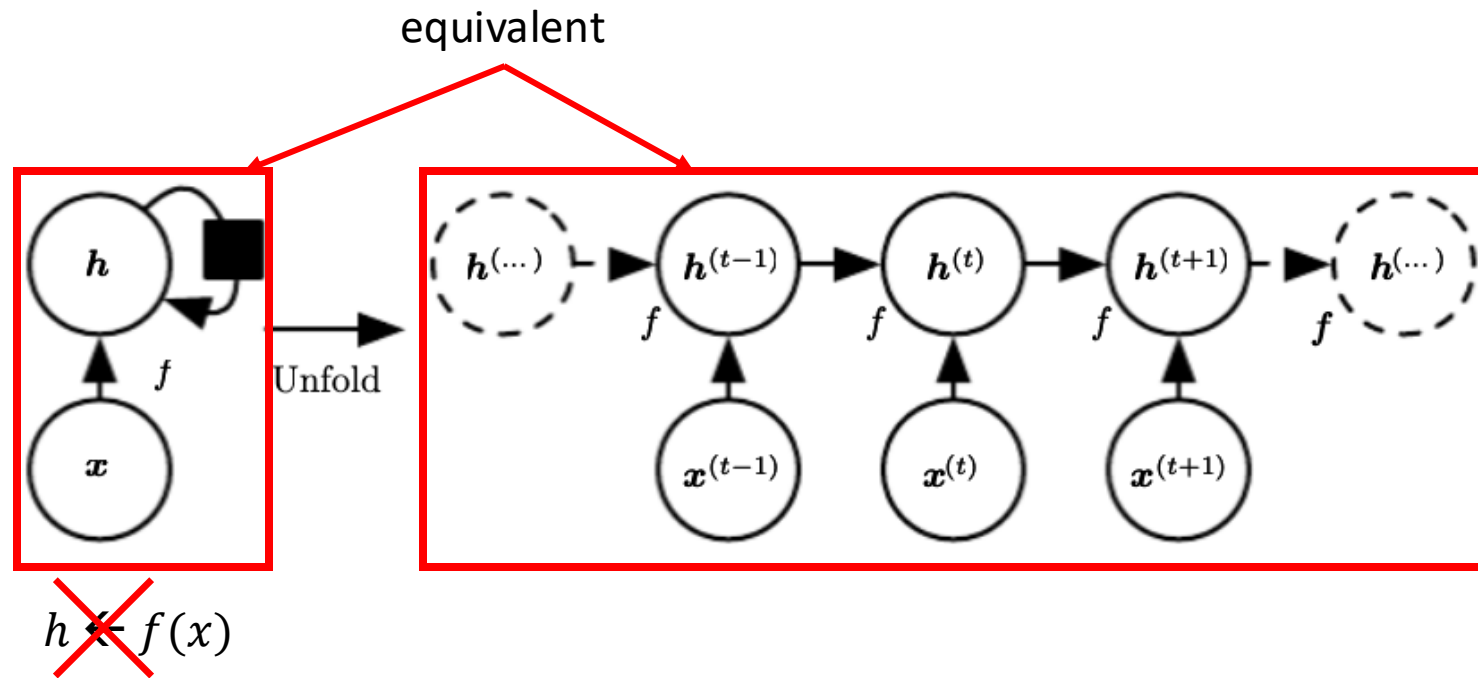
Recurrent networks



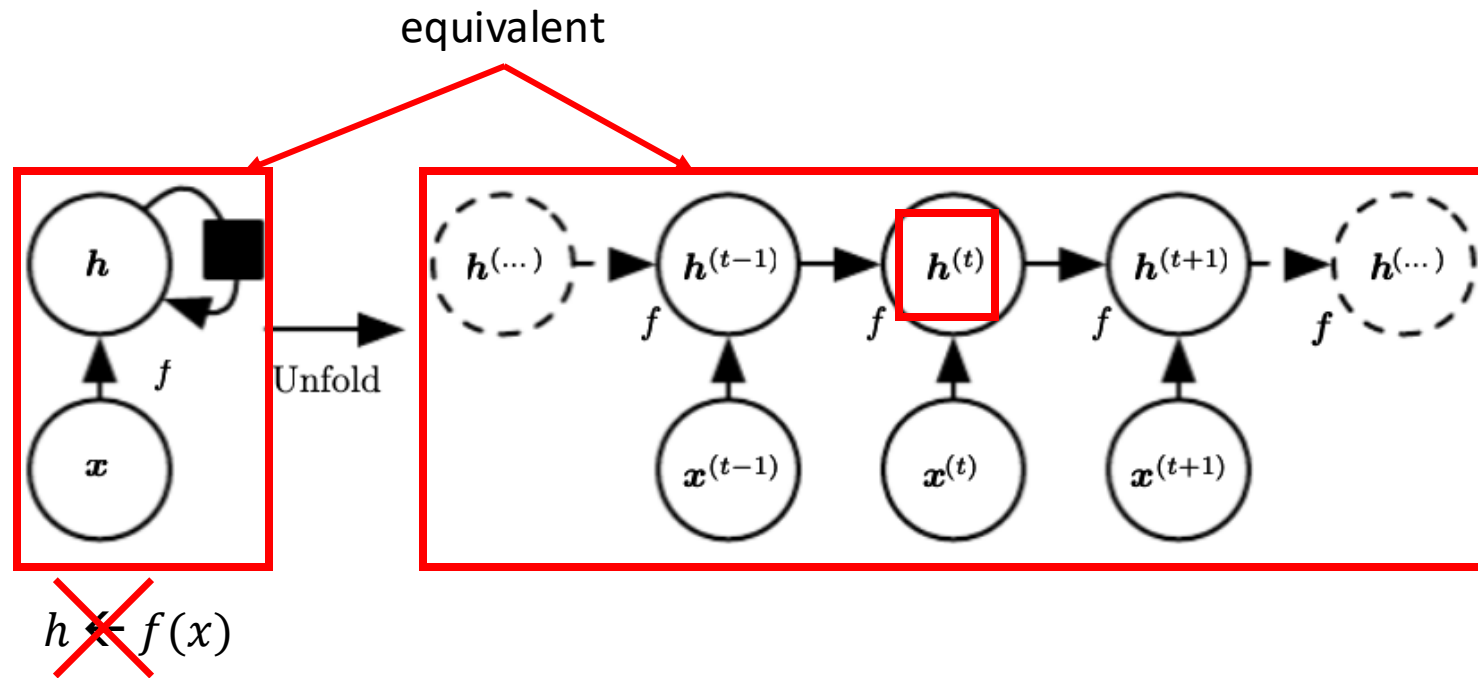
Recurrent networks



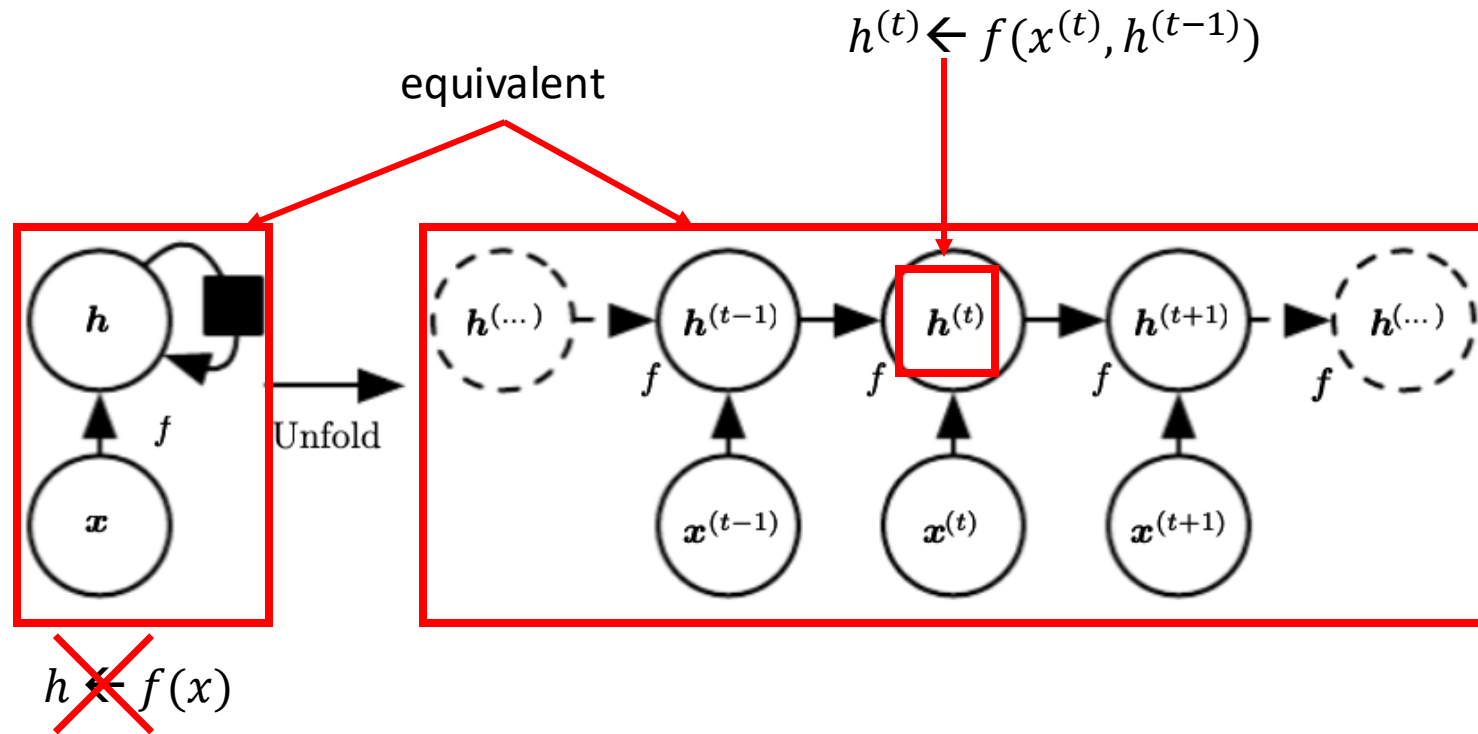
Recurrent networks



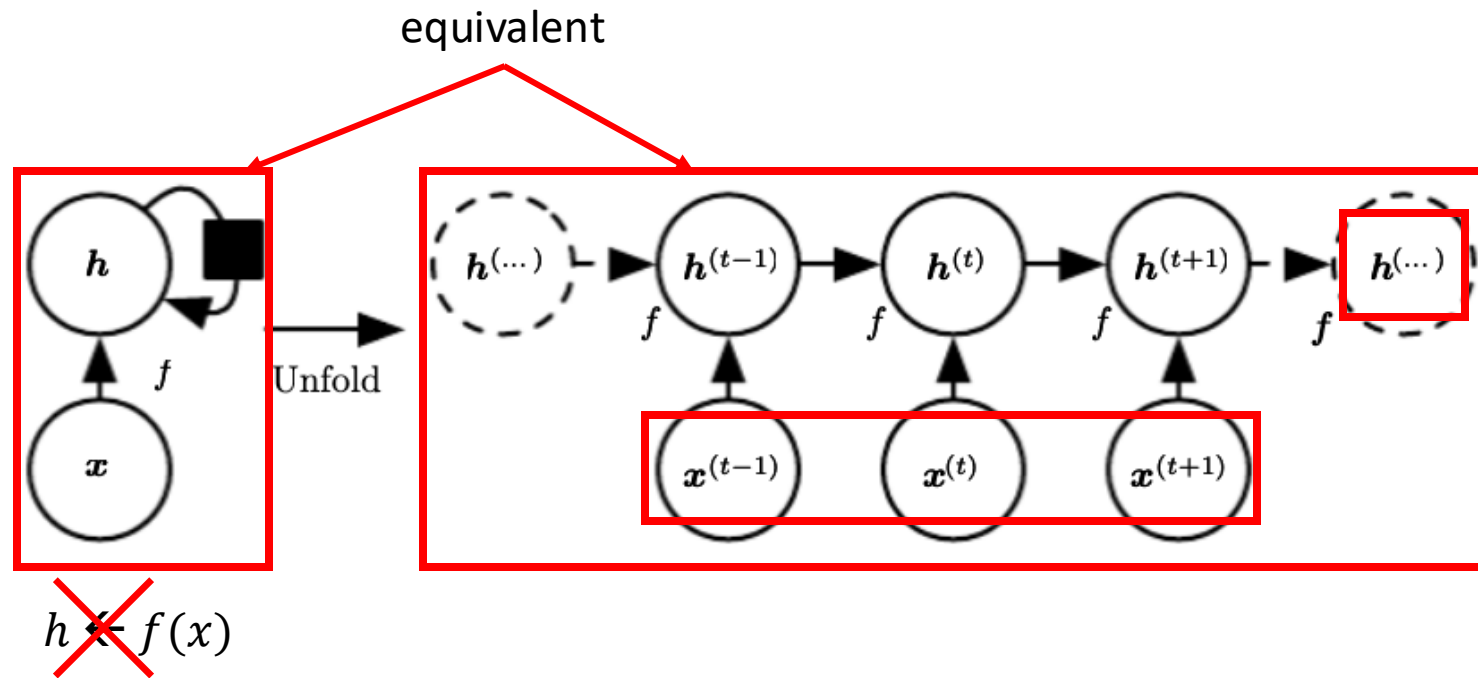
Recurrent networks



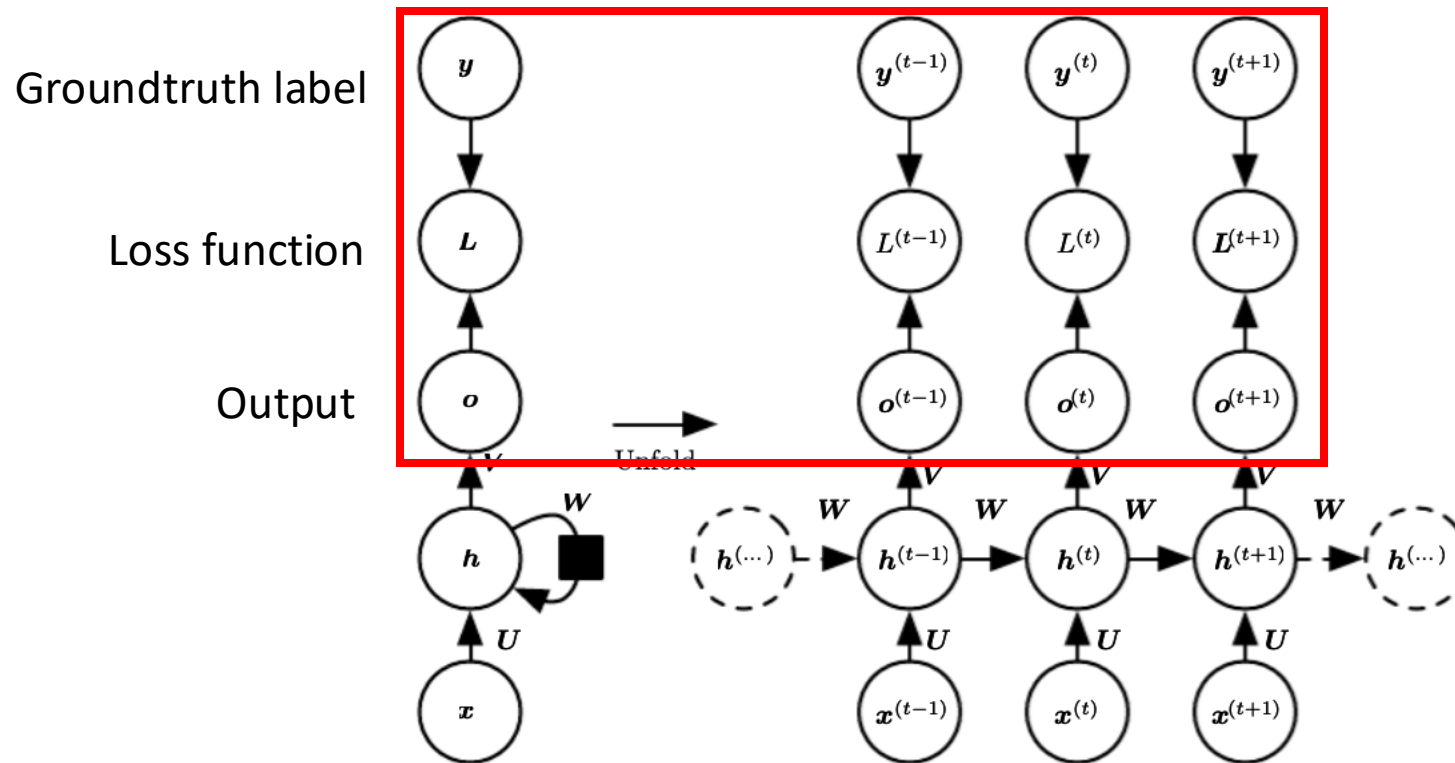
Recurrent networks



Recurrent networks

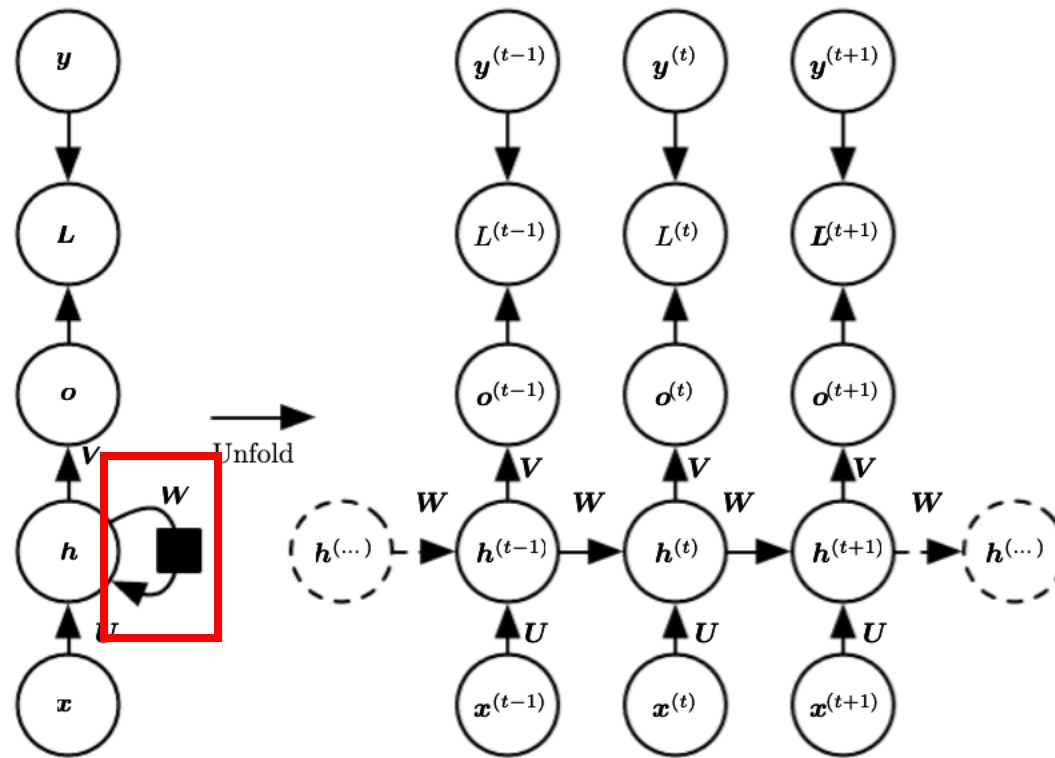


Recurrent networks

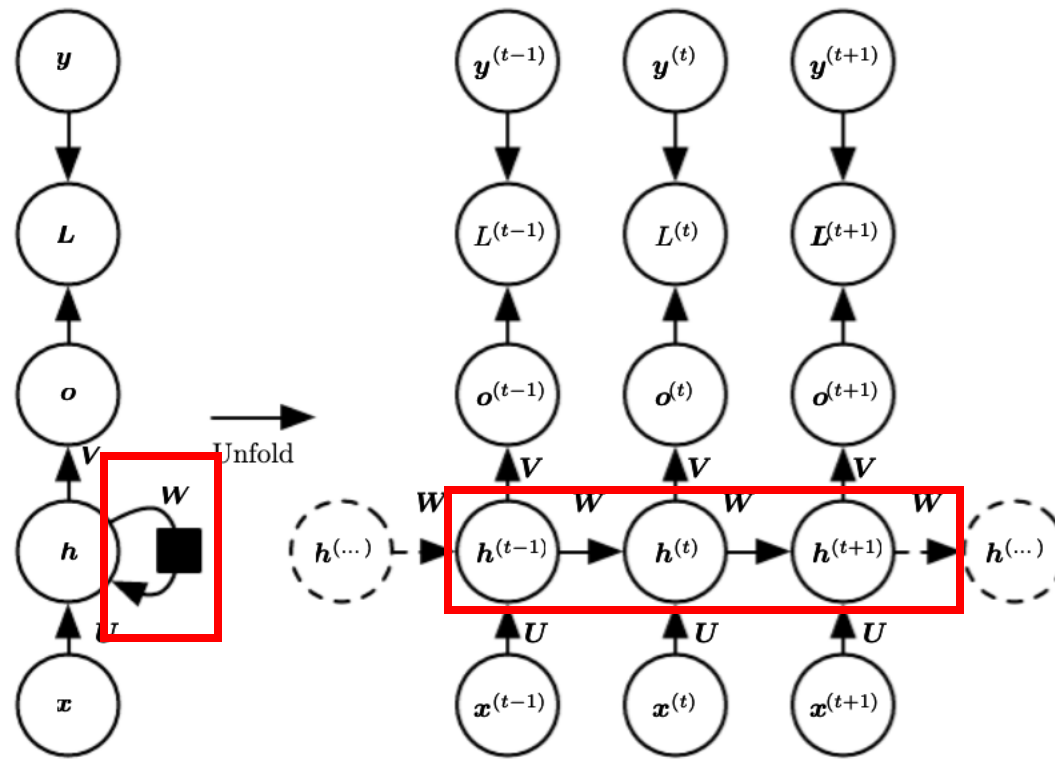


Q: how to describe this structure?

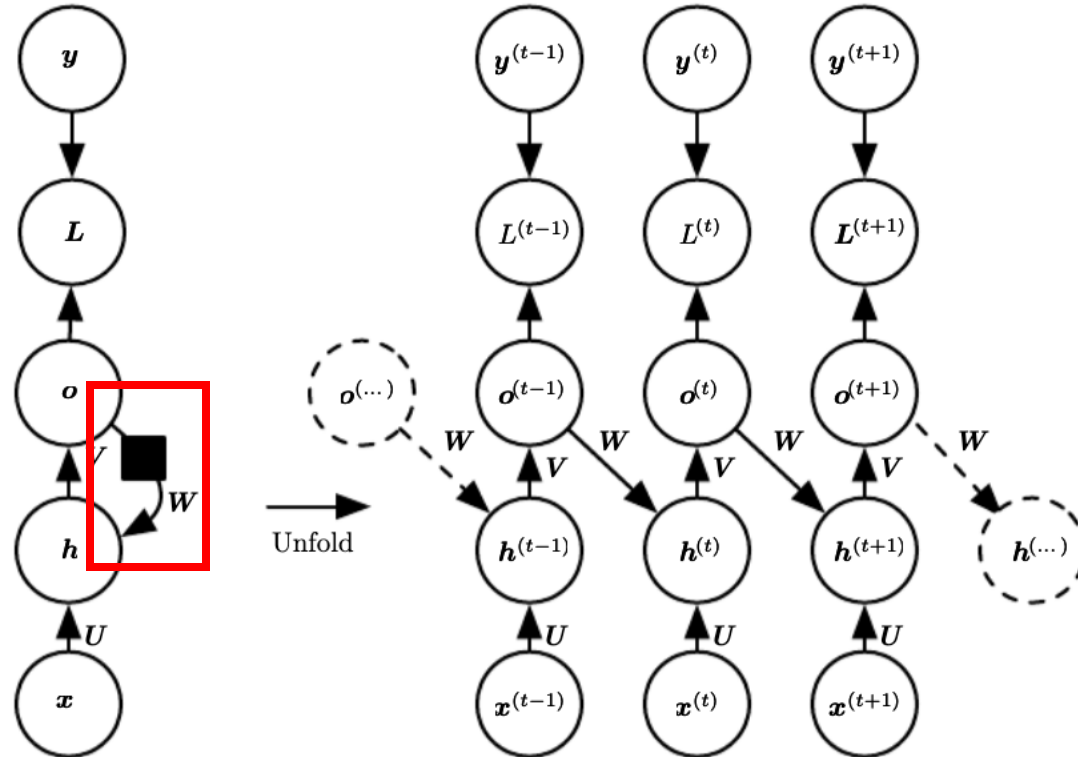
Recurrent networks



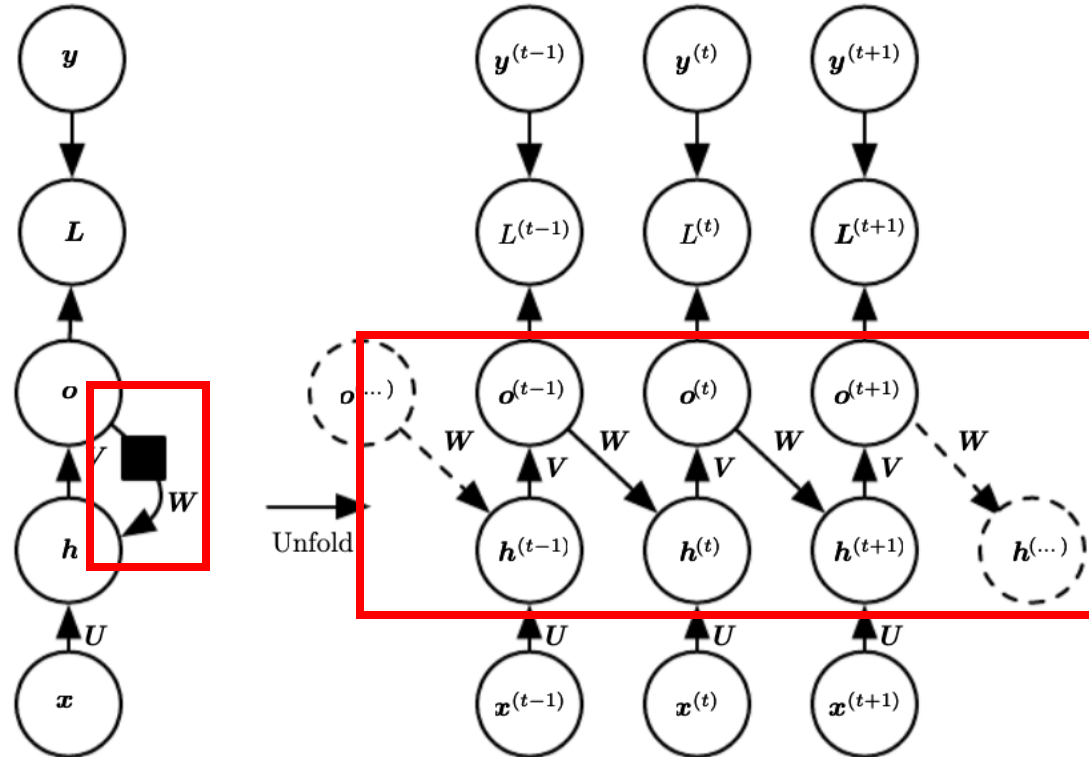
Recurrent networks



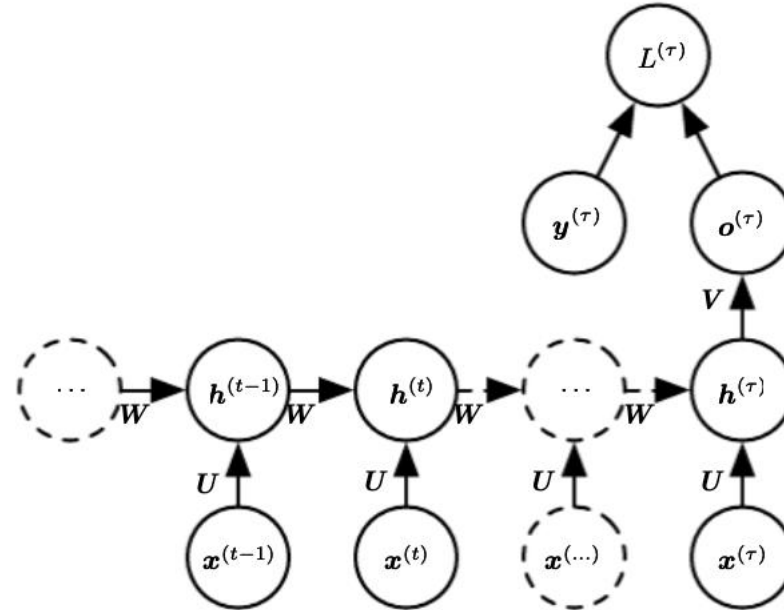
Recurrent networks



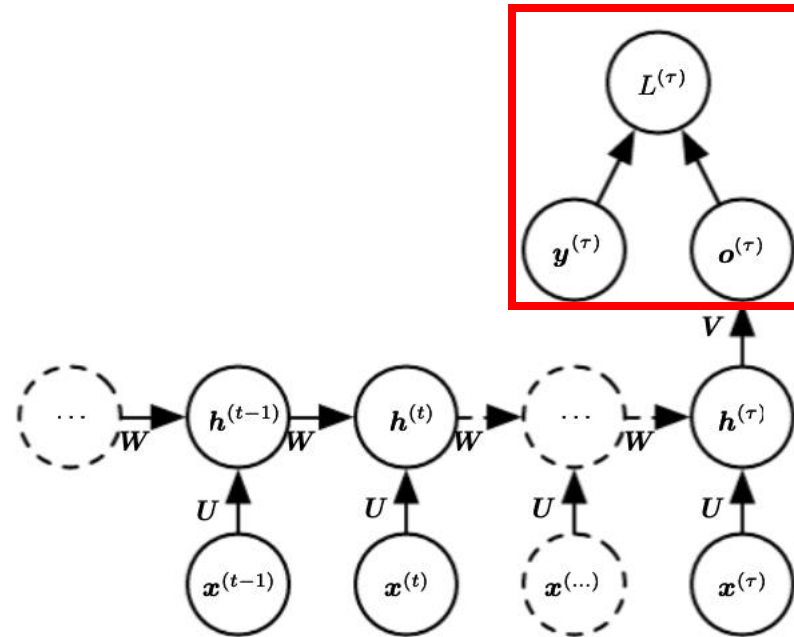
Recurrent networks



Recurrent networks



Recurrent networks

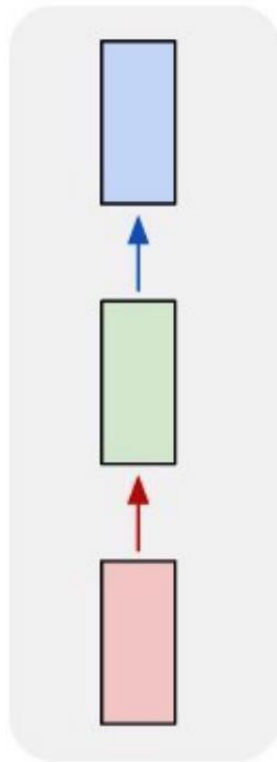


Only one output: summary of a sequence
(Predict a label for a video)

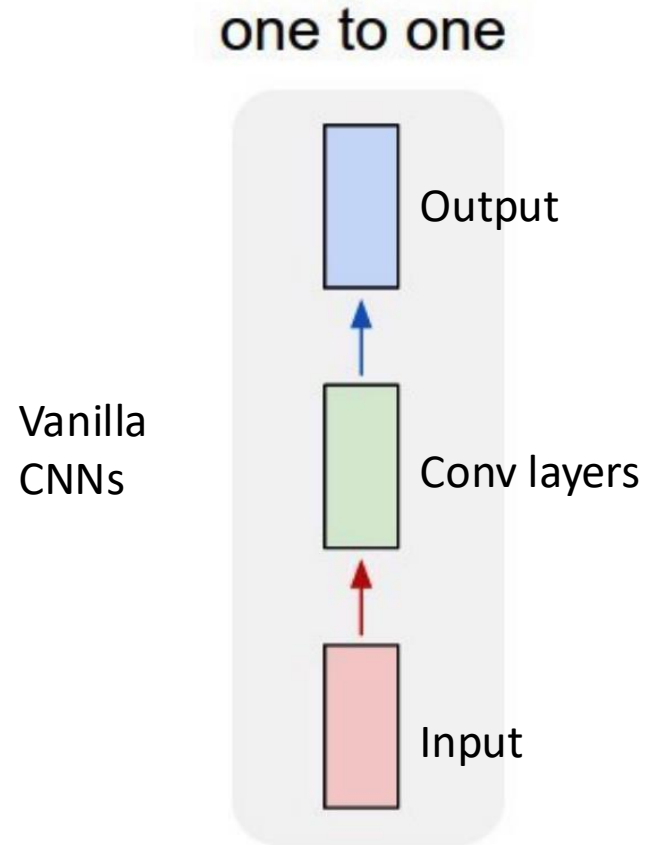
Recurrent neural networks

one to one

Vanilla
CNNs



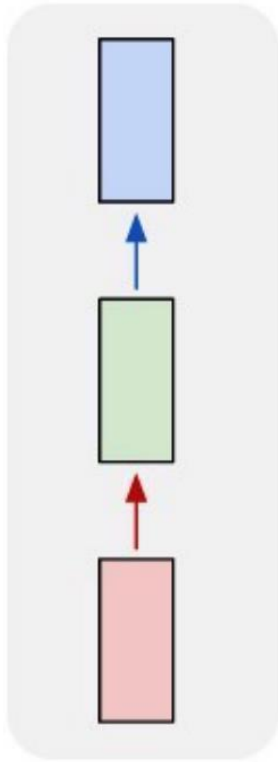
Recurrent neural networks



Recurrent neural networks

one to one

Vanilla
CNNs

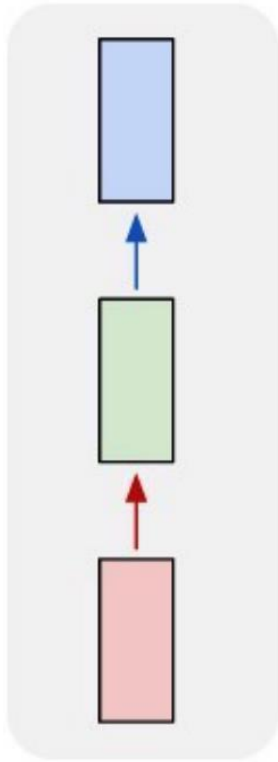


Video data: multiple frames per second

Recurrent neural networks

one to one

Vanilla
CNNs



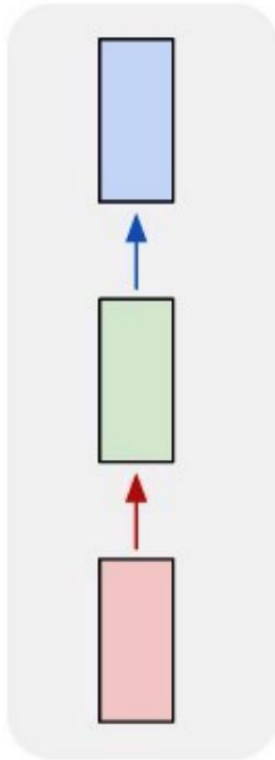
Video data: multiple frames per second

Action recognition

Recurrent neural networks

one to one

Vanilla
CNNs

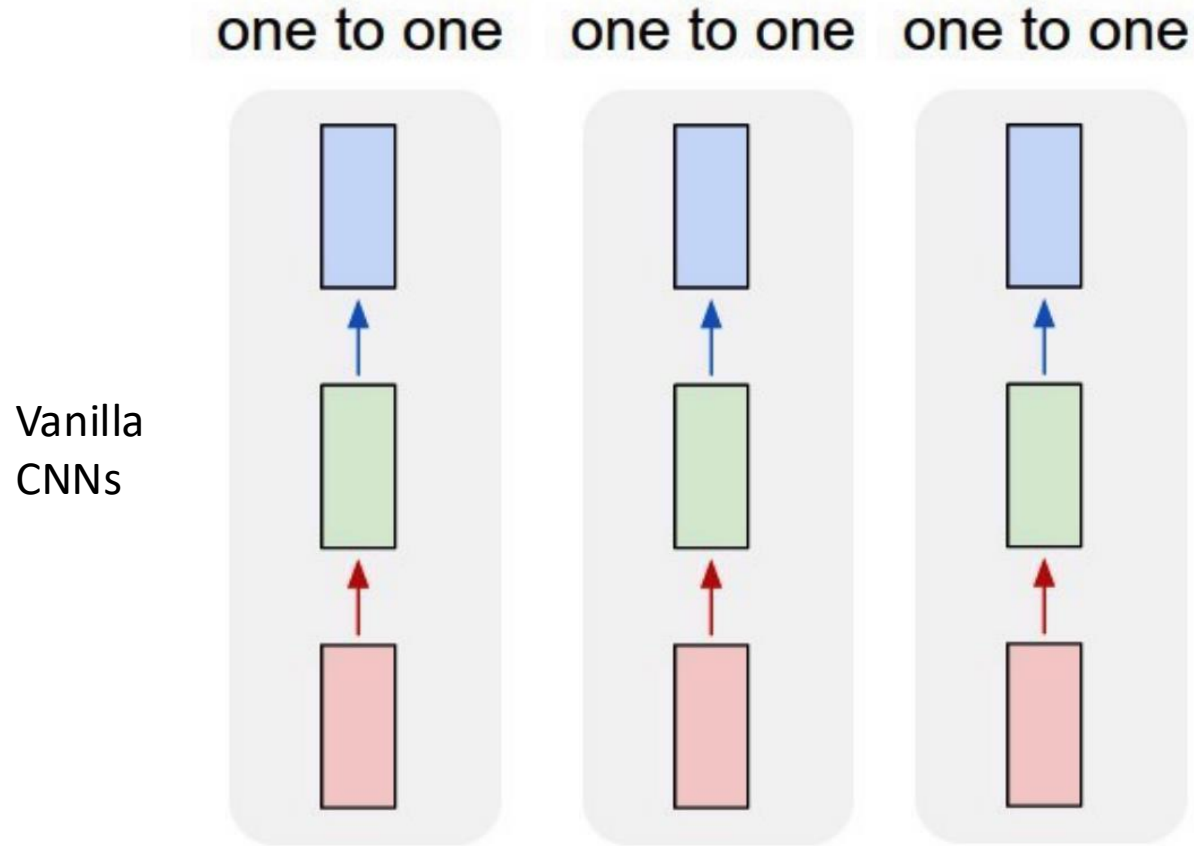


Video data: multiple frames per second

Action recognition



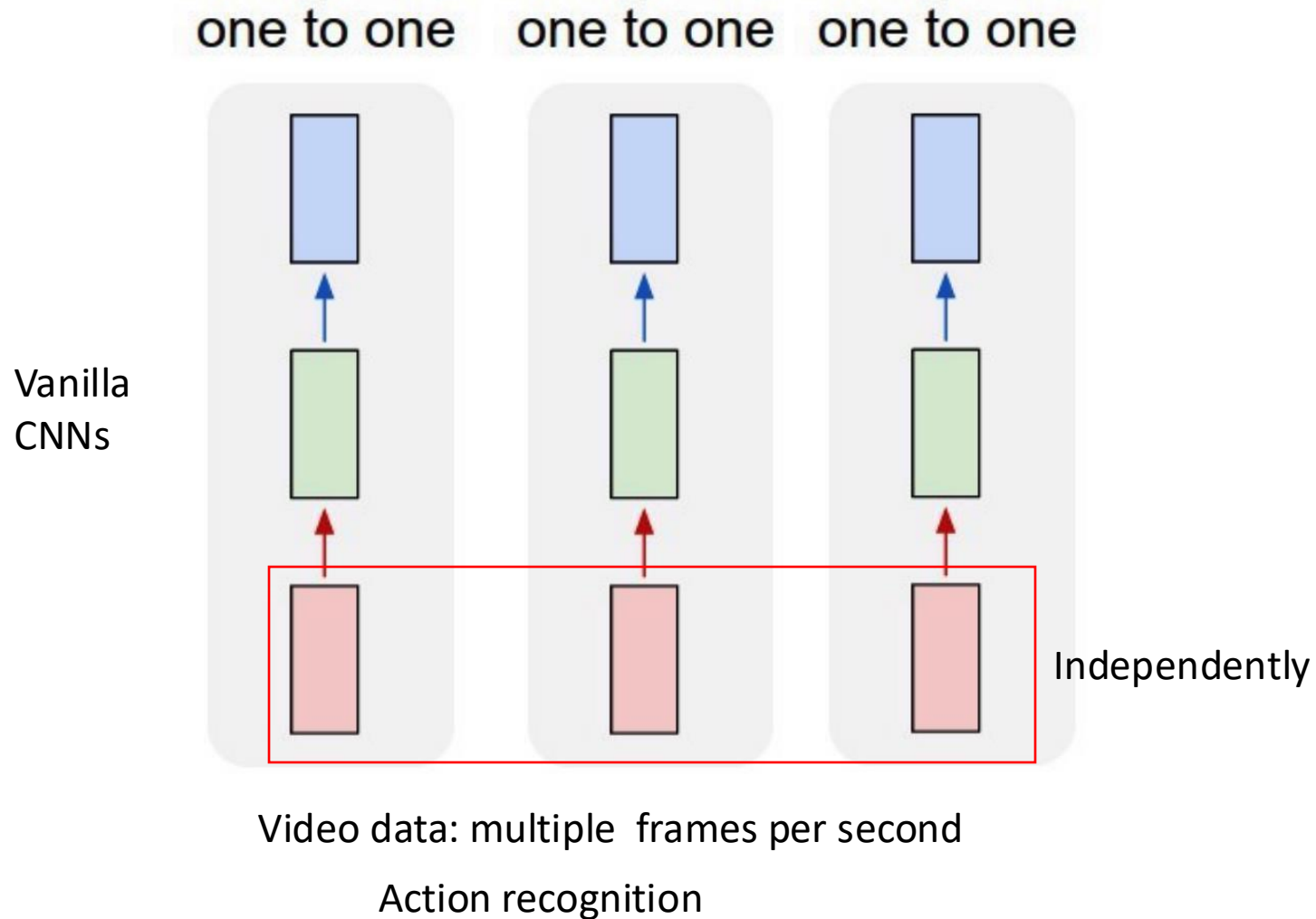
Recurrent neural networks



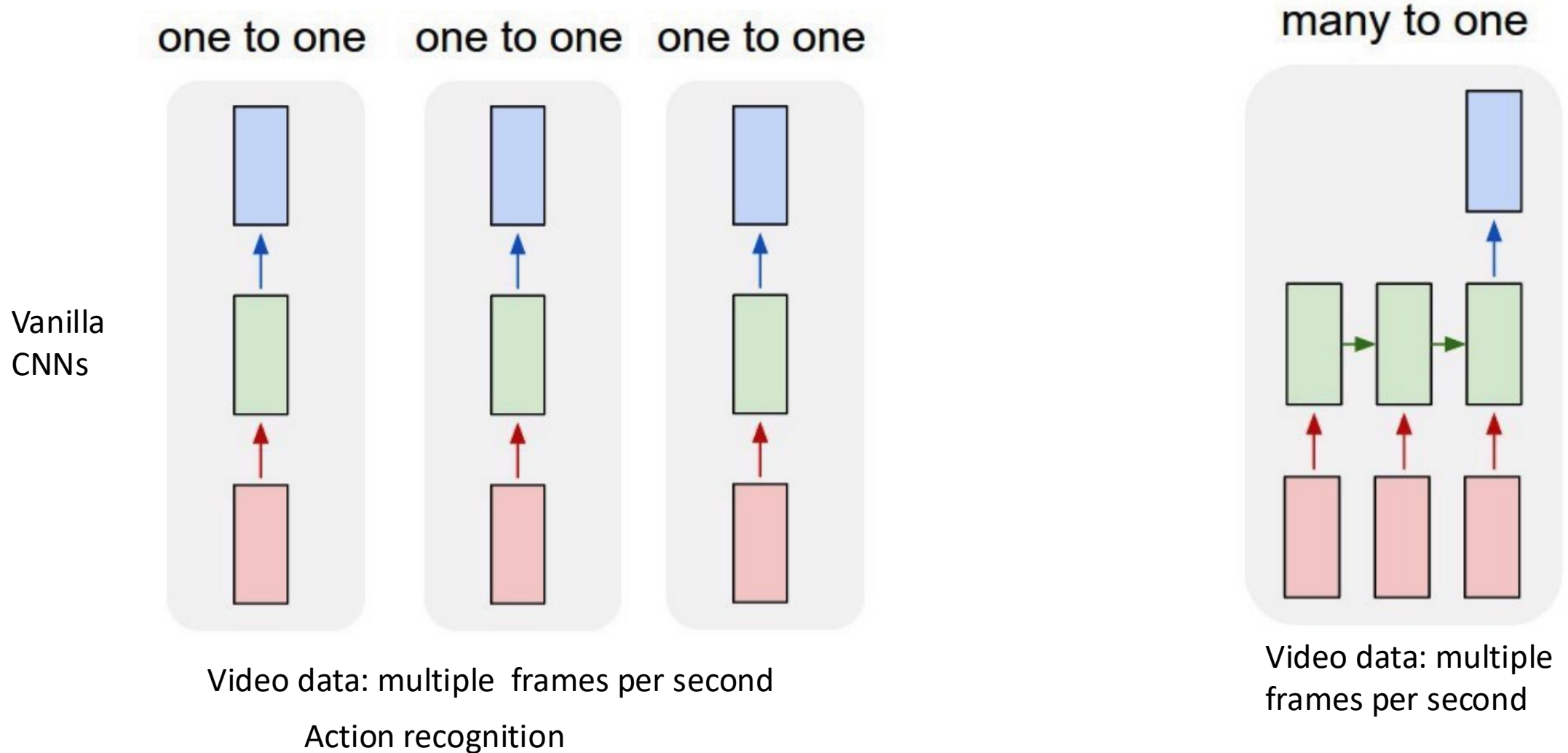
Video data: multiple frames per second

Action recognition

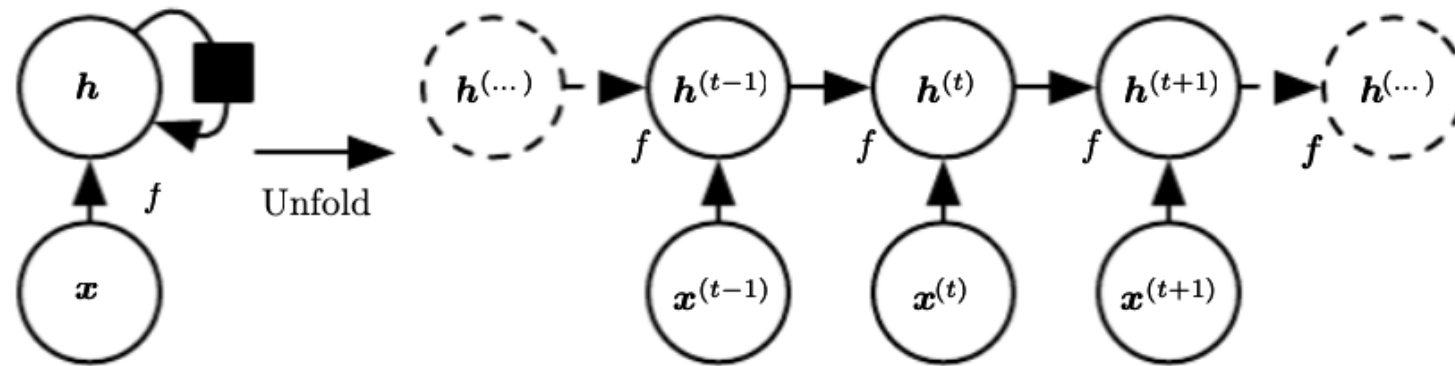
Recurrent neural networks



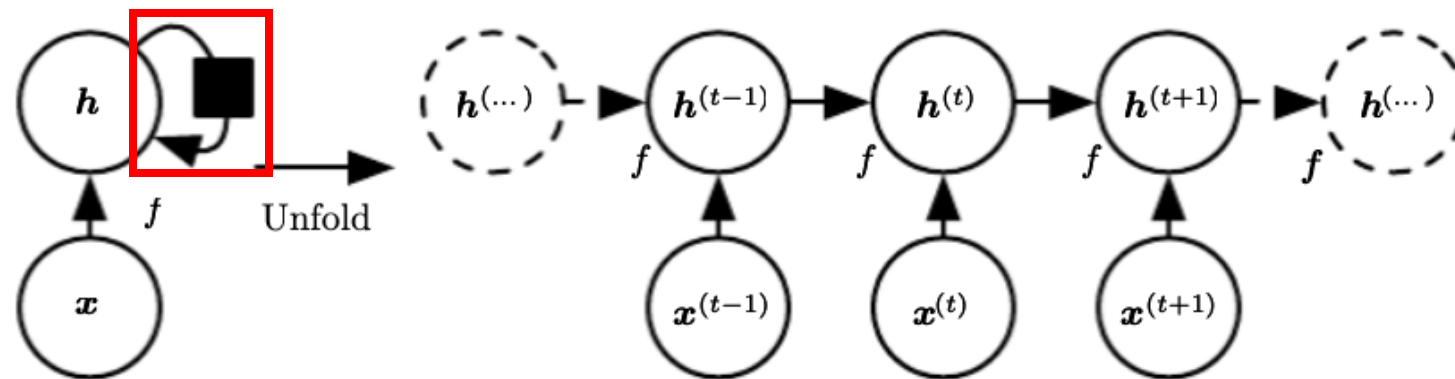
Recurrent neural networks



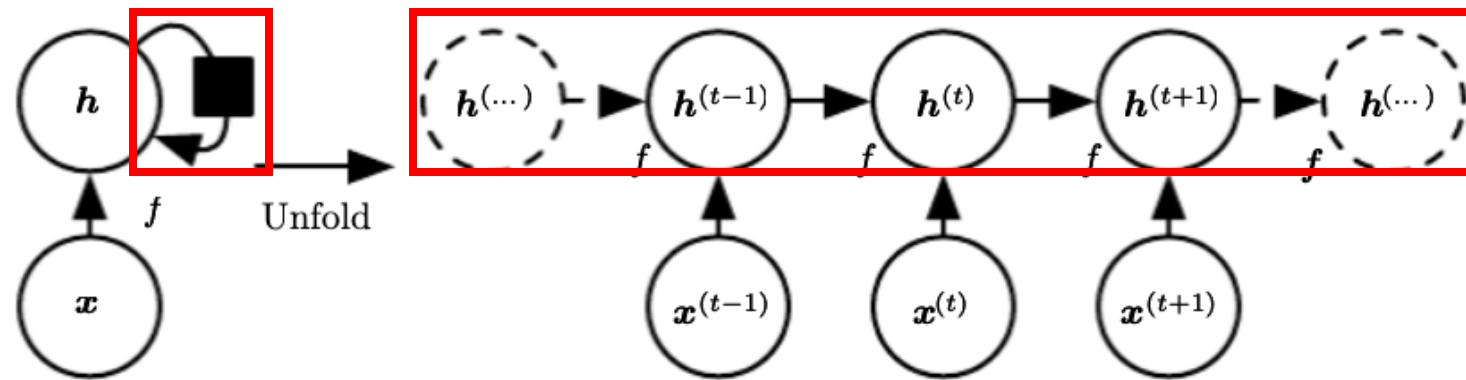
Recurrent neural networks



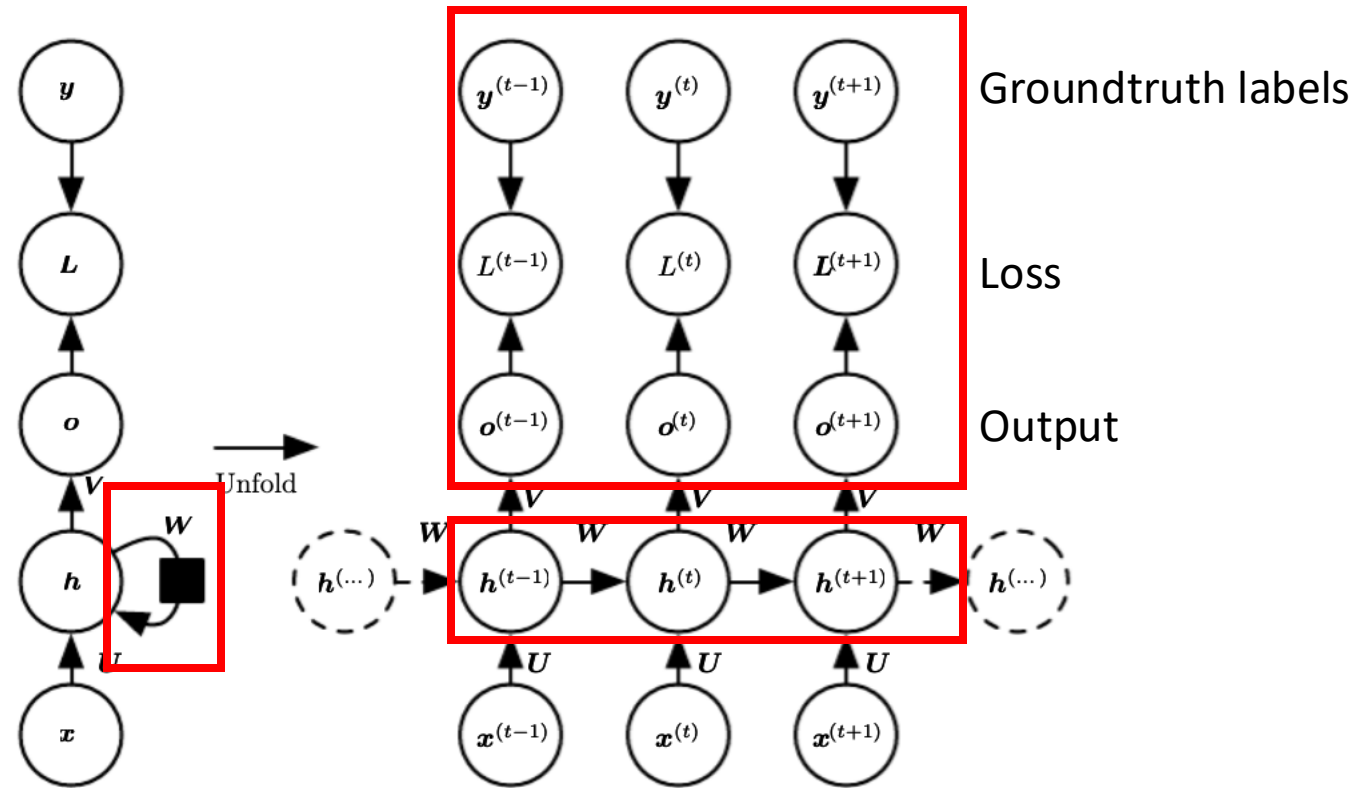
Recurrent neural networks



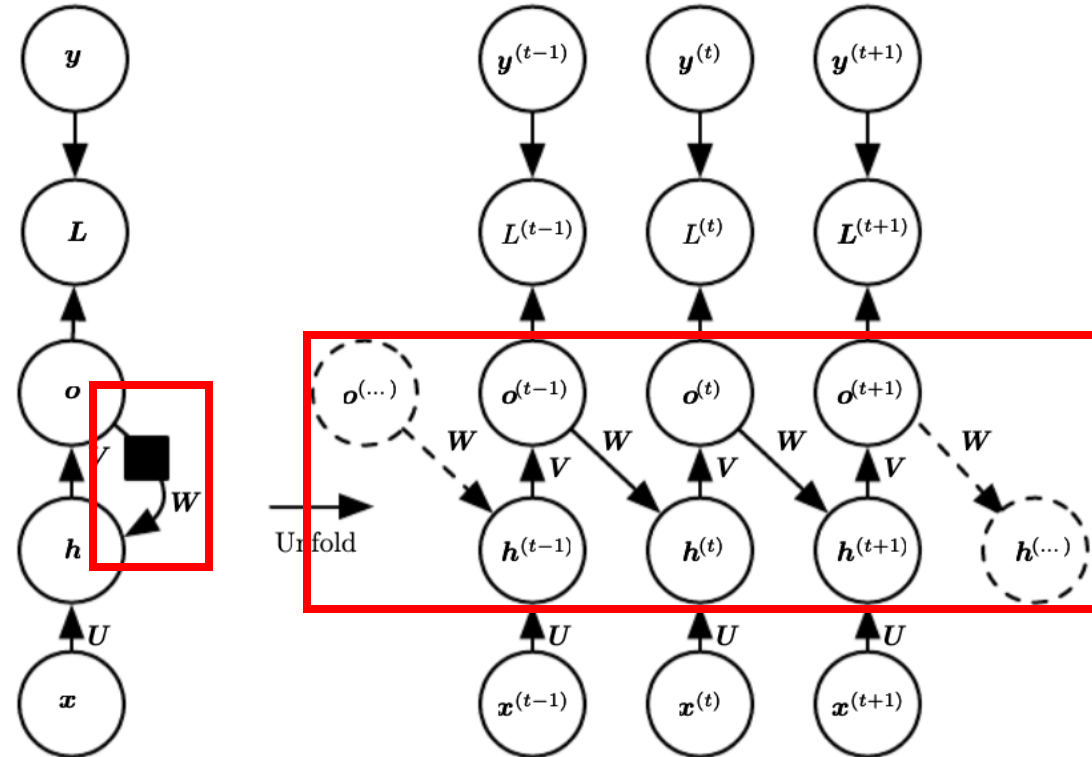
Recurrent neural networks



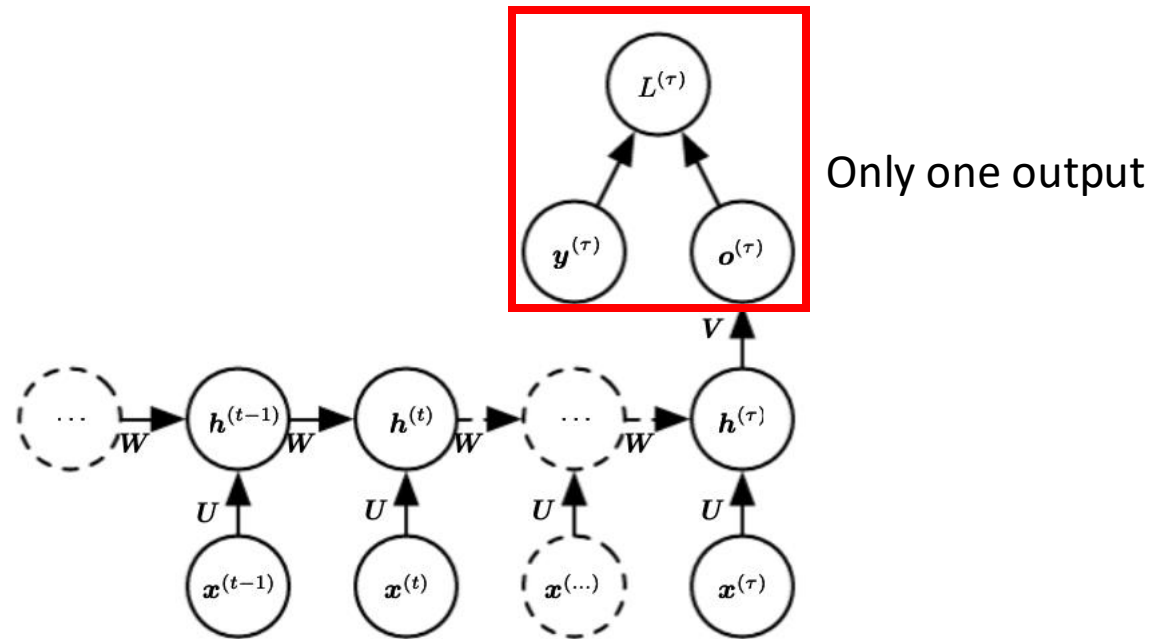
Recurrent networks



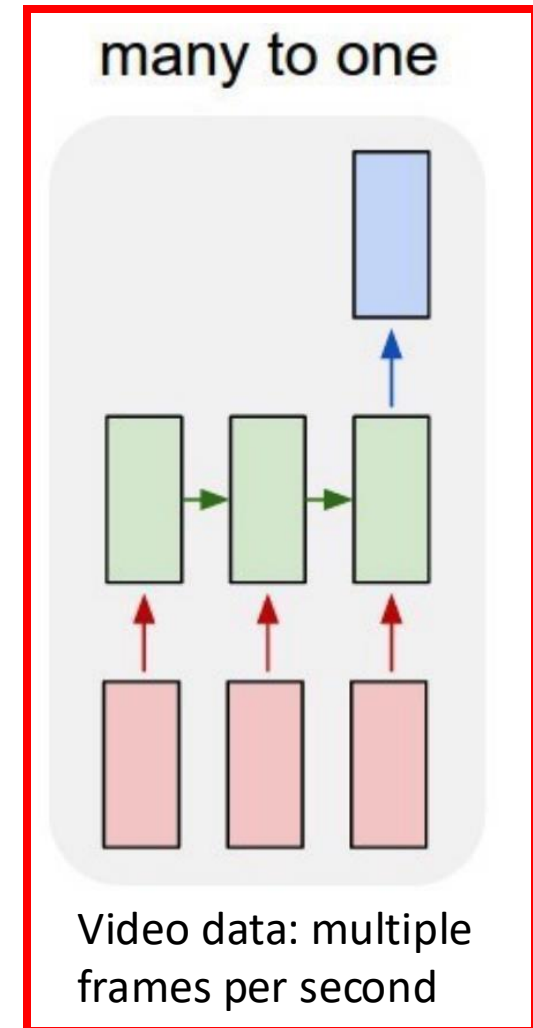
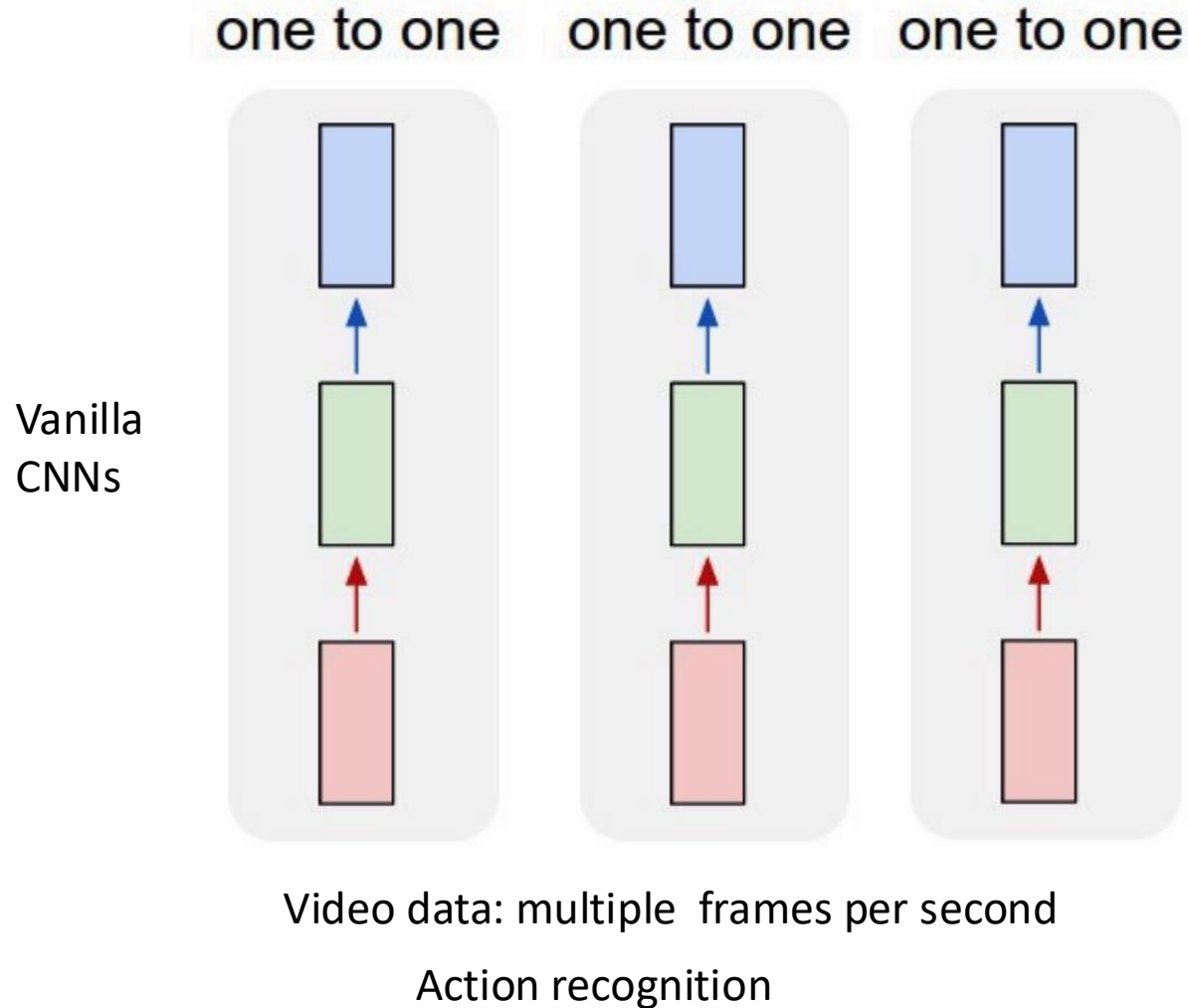
Recurrent networks



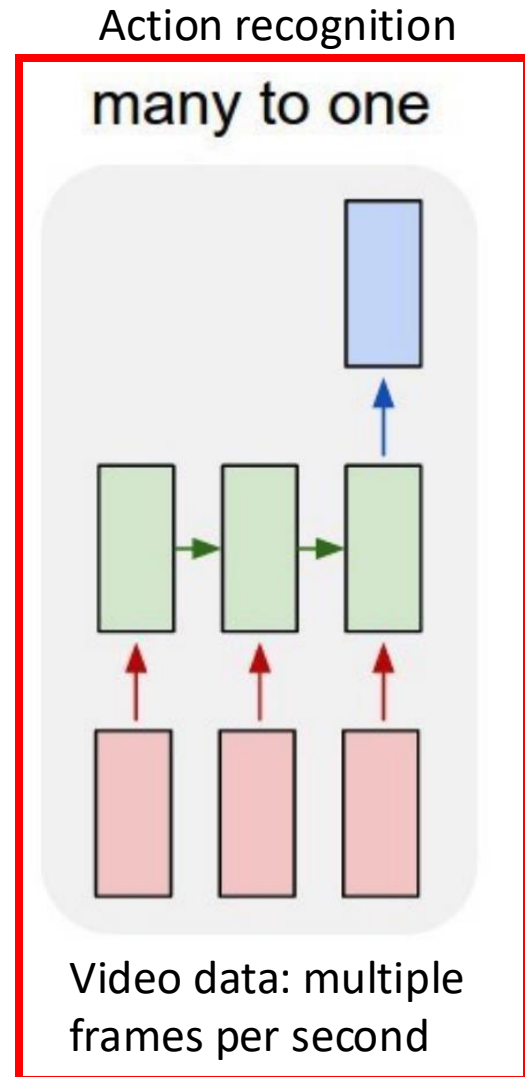
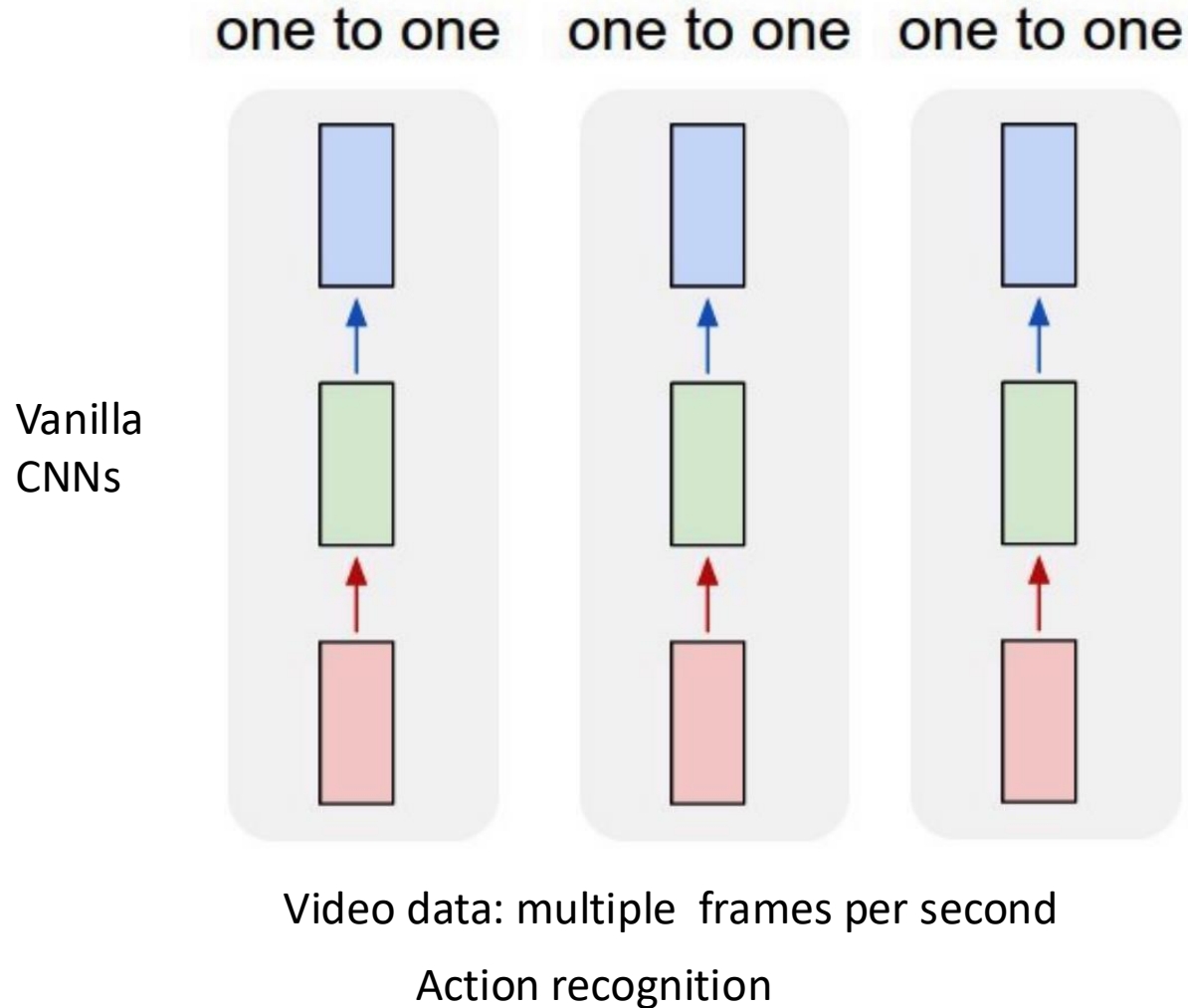
Recurrent networks



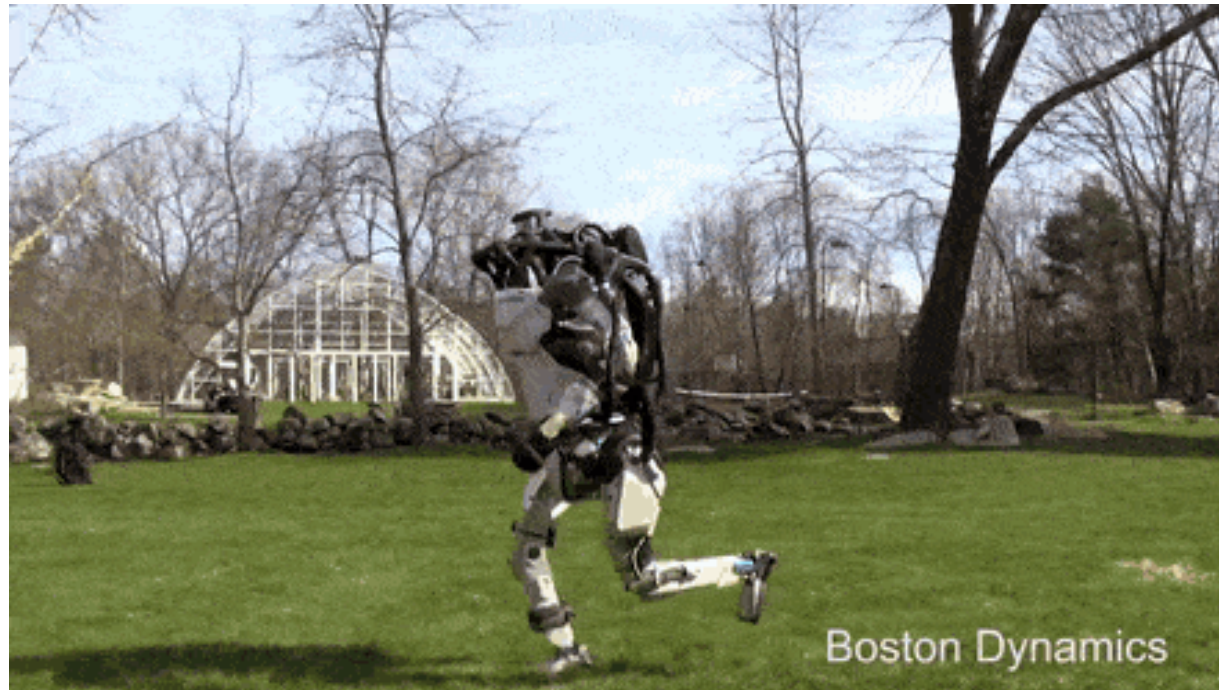
Recurrent neural networks



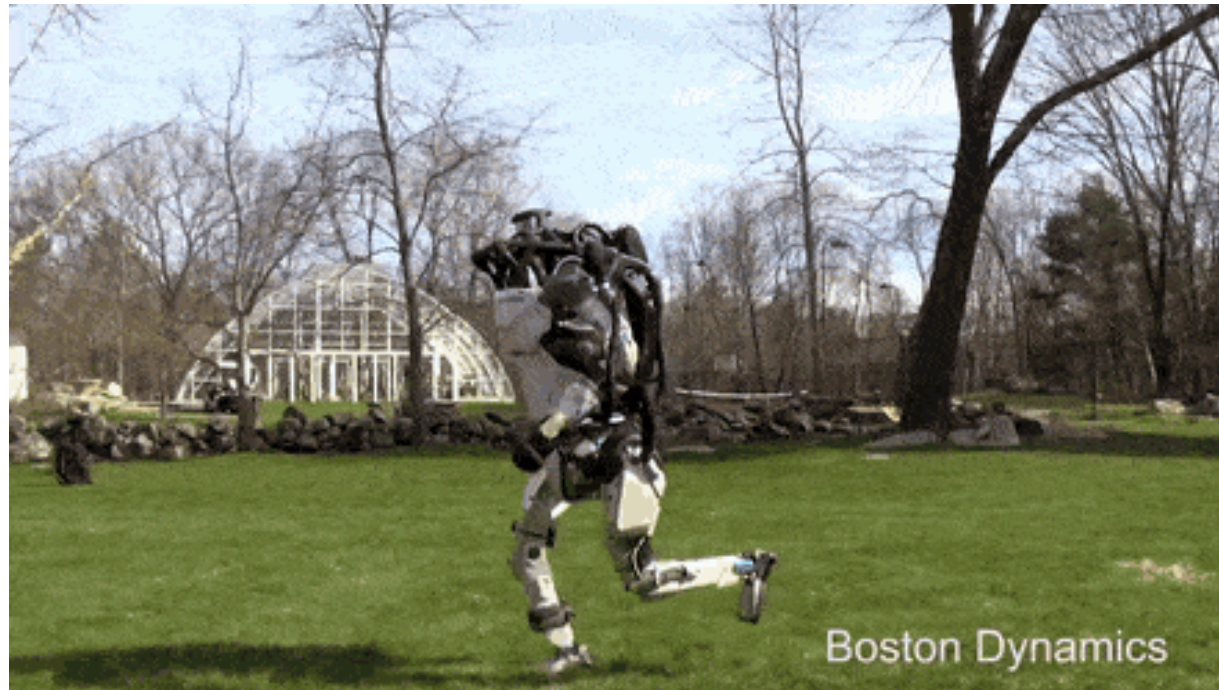
Recurrent neural networks



Recurrent neural networks in practice

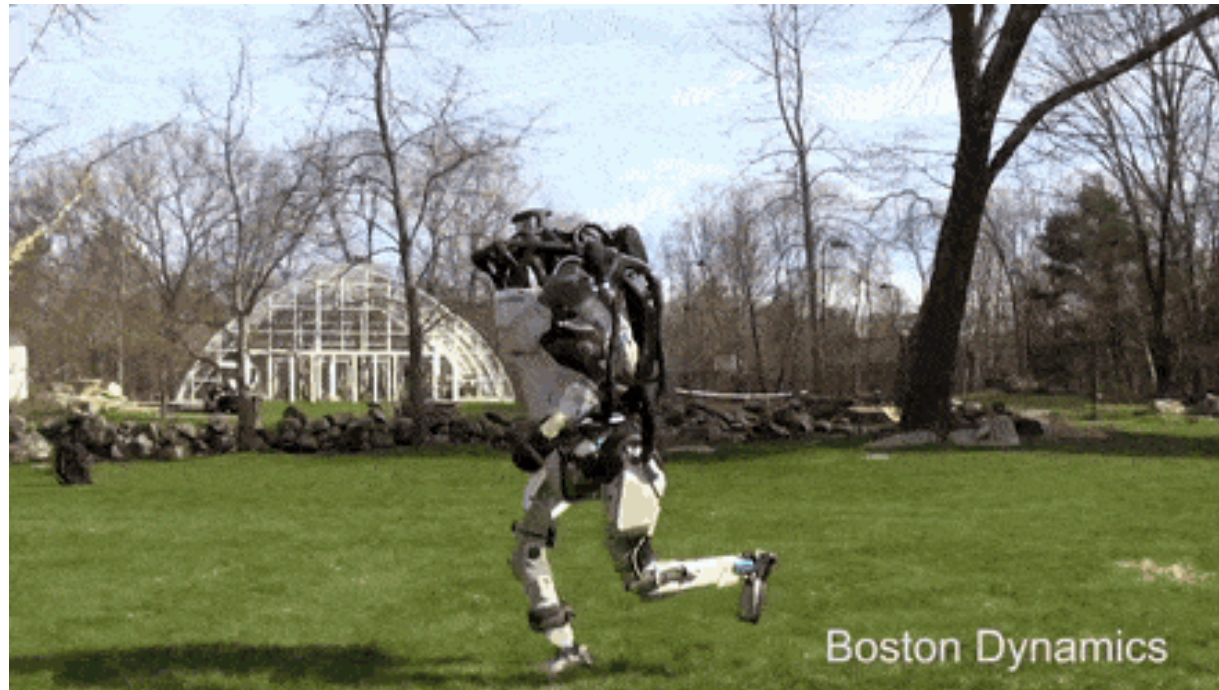


Recurrent neural networks in practice



Q: what is the action?

Recurrent neural networks in practice



Q: what is the action?

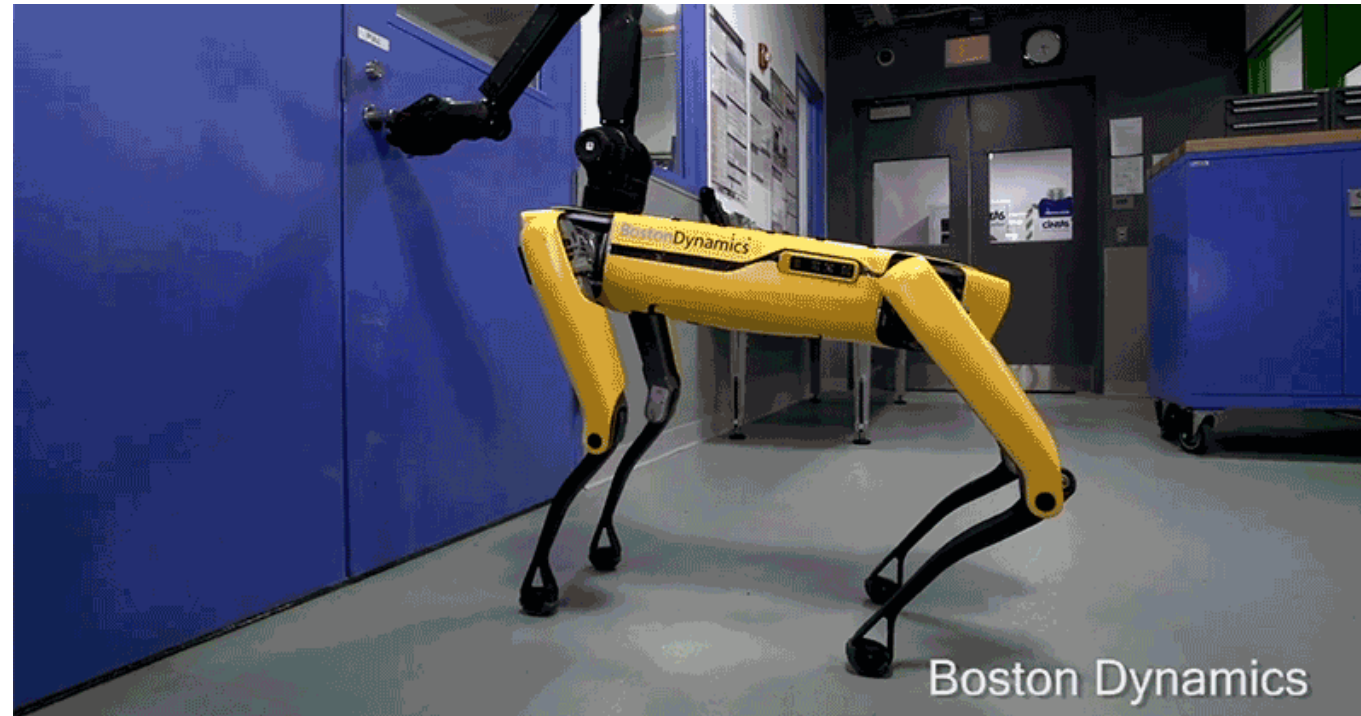
Running or opening a door?

Recurrent neural networks in practice



Q: what is the action?

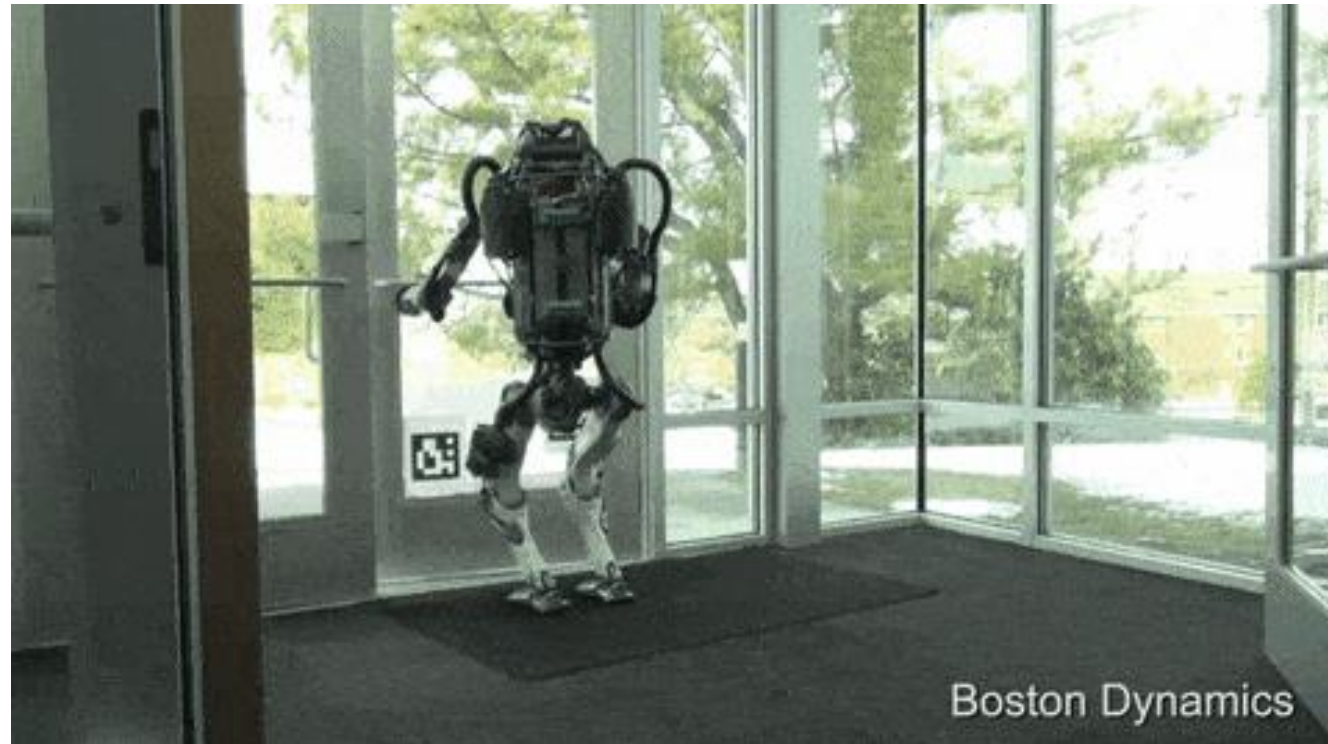
Recurrent neural networks in practice



Q: what is the action?

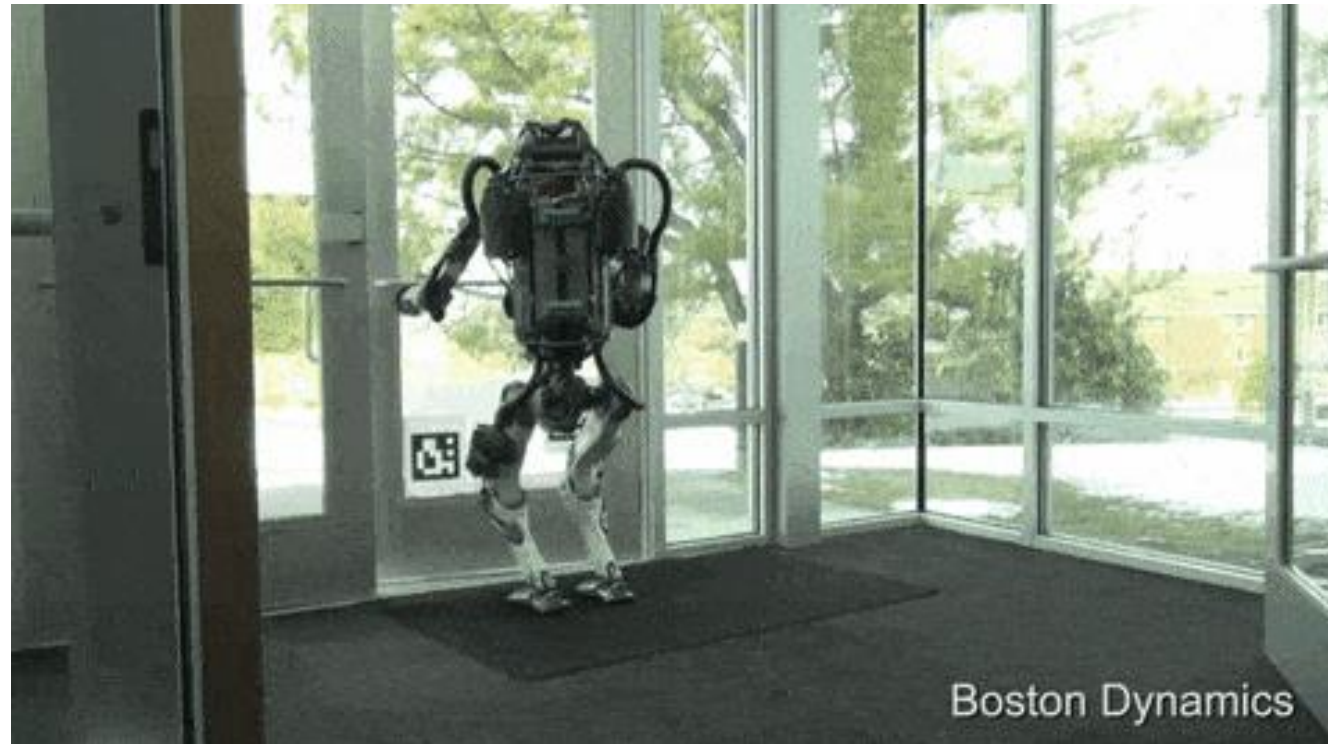
Running or opening a door?

Recurrent neural networks in practice



Q: what is the action?

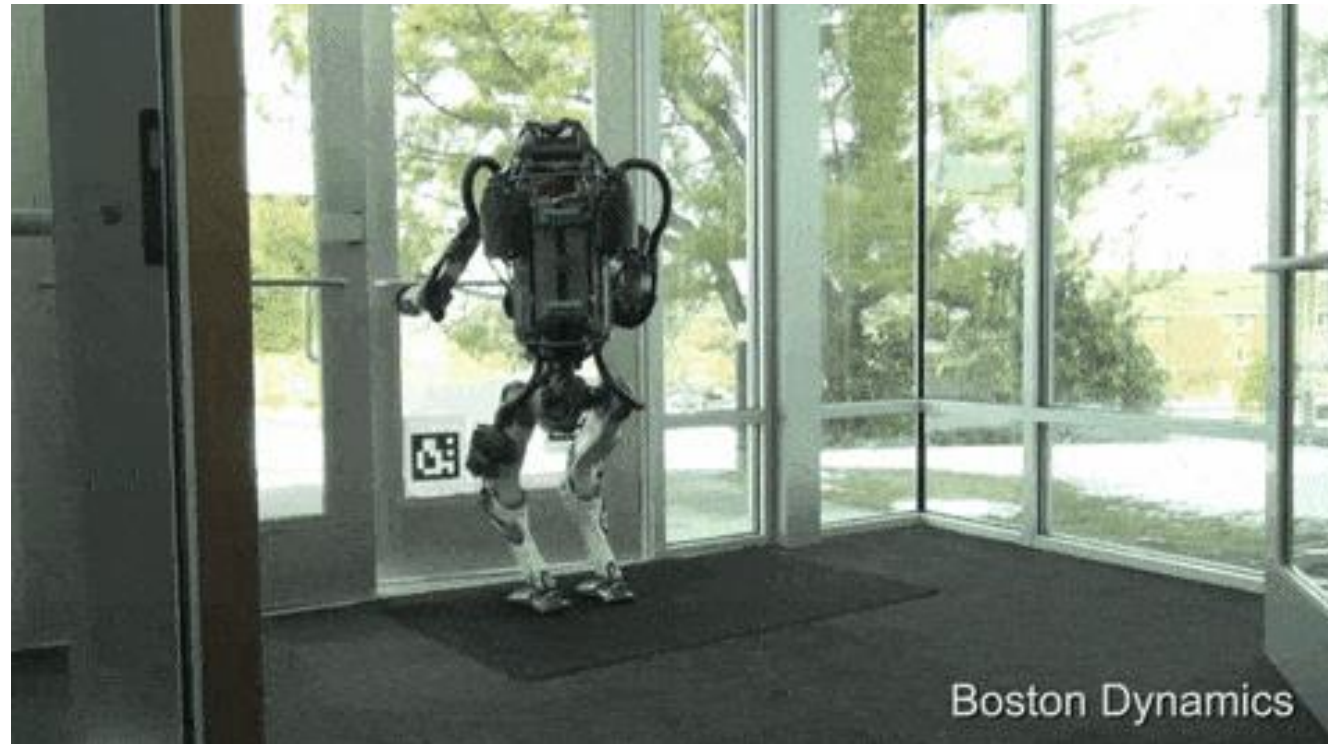
Recurrent neural networks in practice



Q: what is the action?

Running or opening a door?

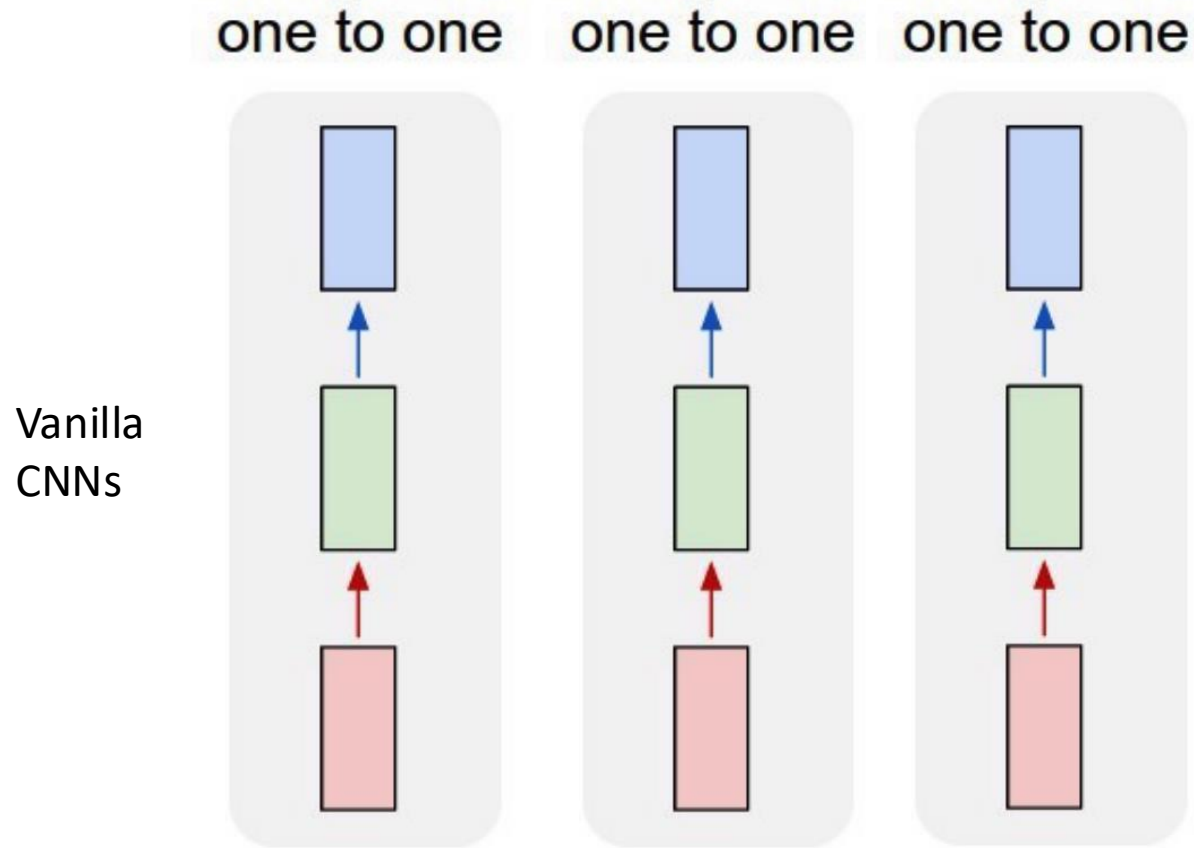
Recurrent neural networks in practice



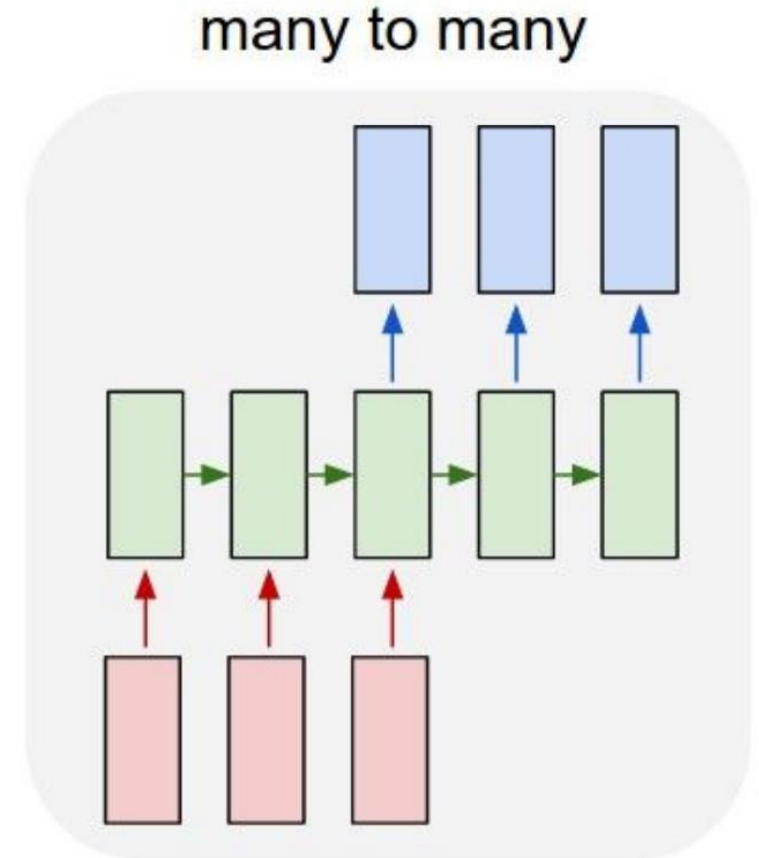
Action recognition:
predict a label from given multiple frames

Q: what is the action?
Running or opening a door?

Recurrent neural networks



Video data: multiple frames per second
Action recognition

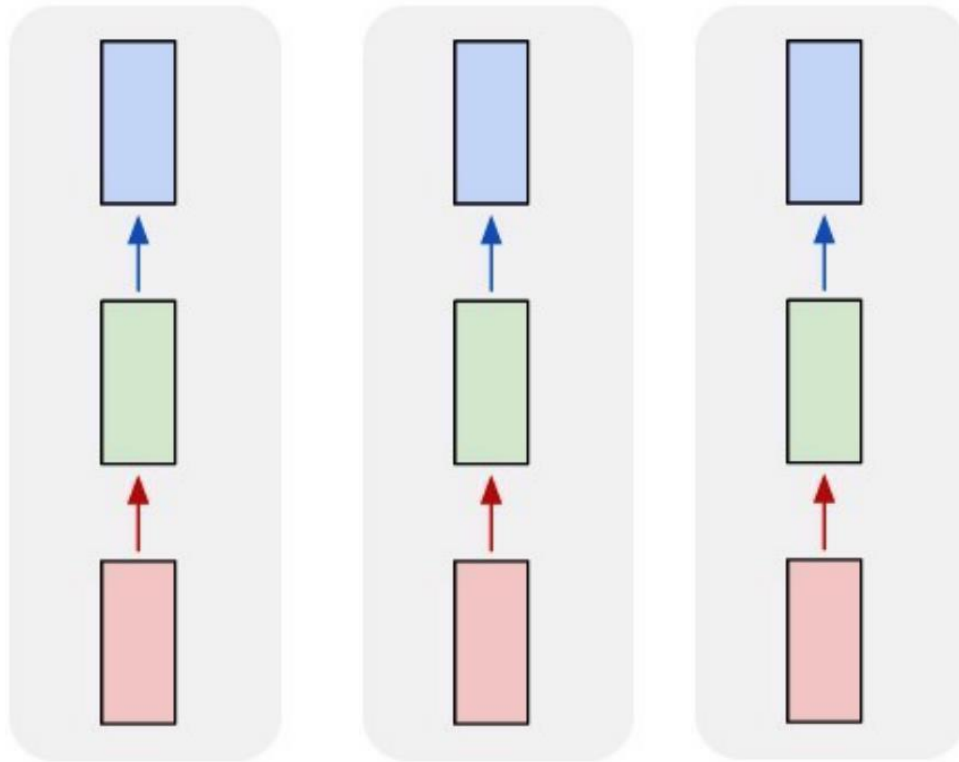


Video data: multiple frames per second

Recurrent neural networks

Vanilla
CNNs

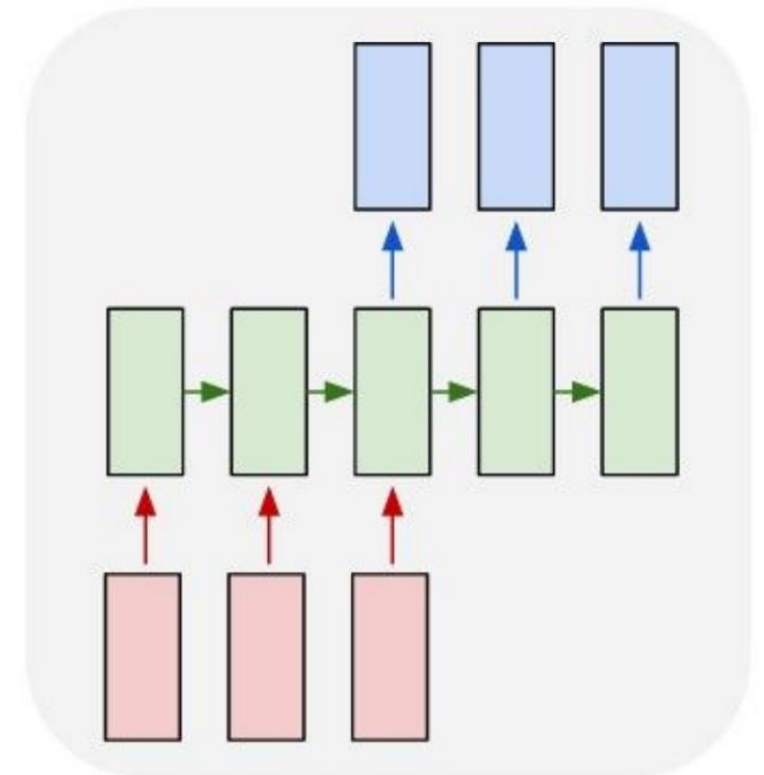
one to one one to one one to one



Video data: multiple frames per second
Action recognition

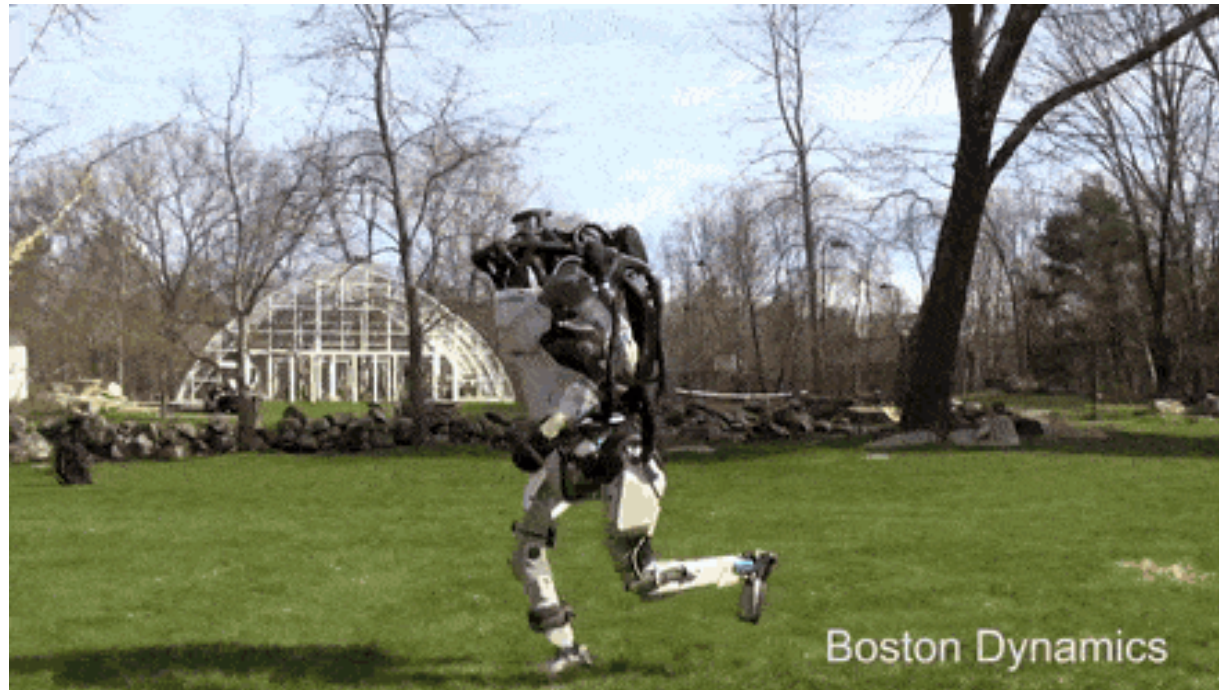
Q: what application?

many to many



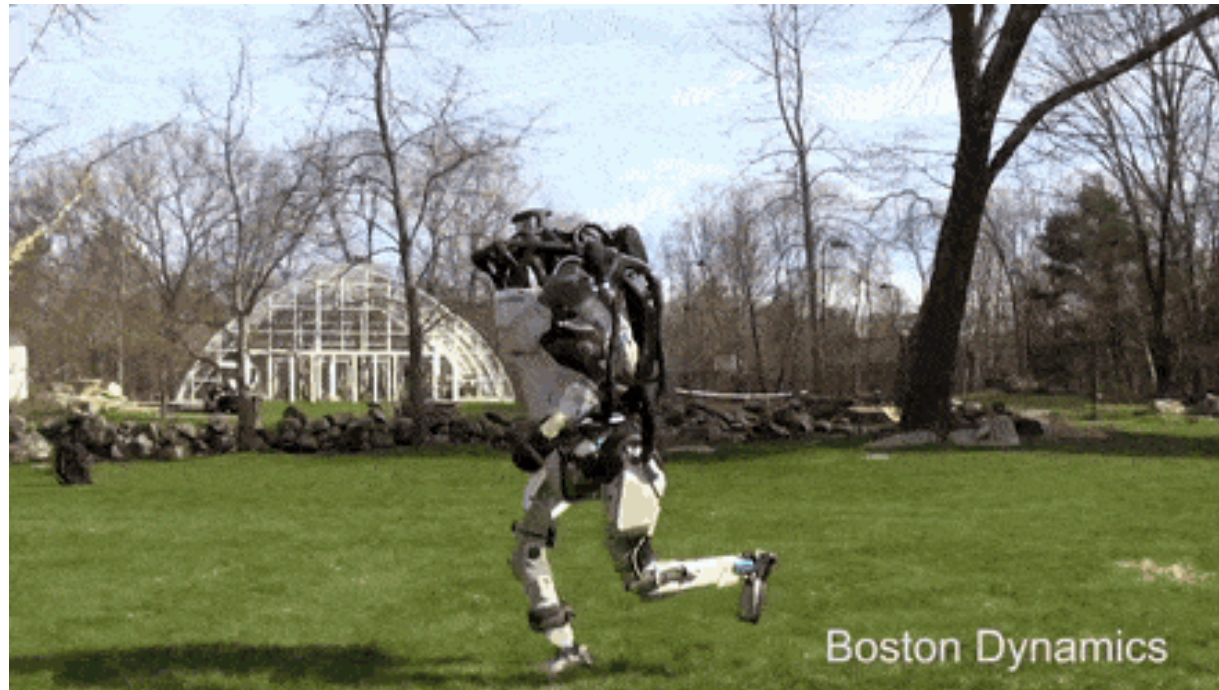
Video data: multiple frames per second

Recurrent neural networks in practice



Q: what is the action?

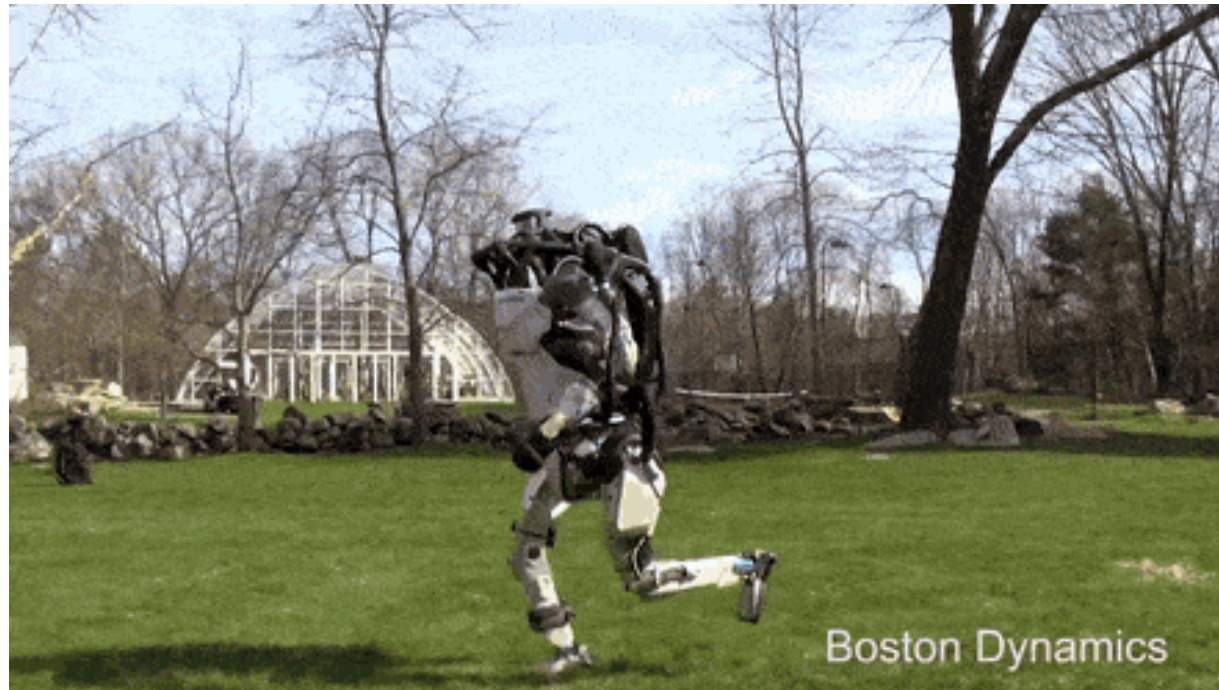
Recurrent neural networks in practice



Q: what is the action?

R

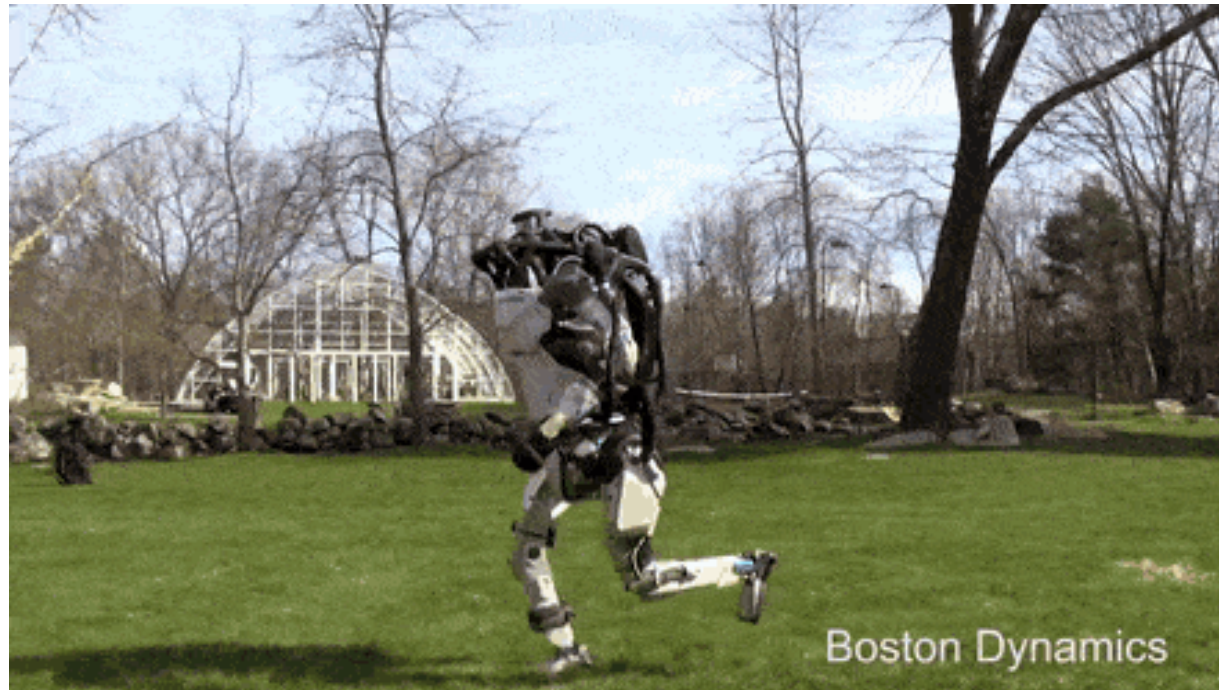
Recurrent neural networks in practice



Q: what is the action?

Ru

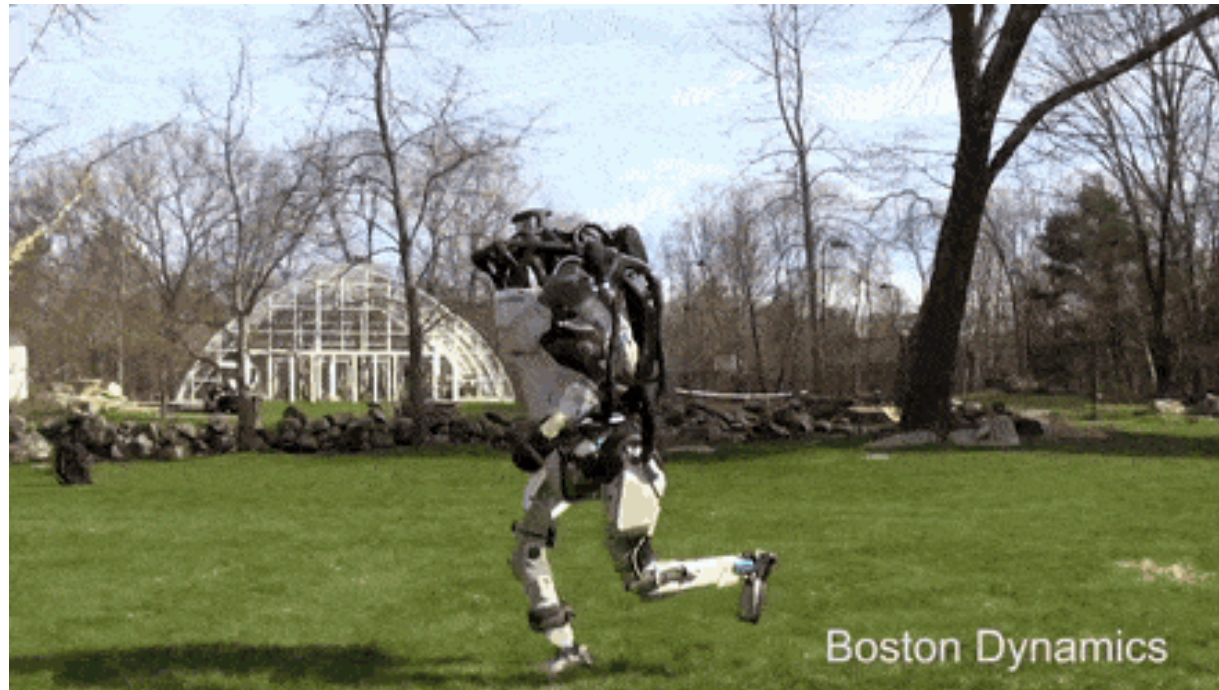
Recurrent neural networks in practice



Q: what is the action?

Run

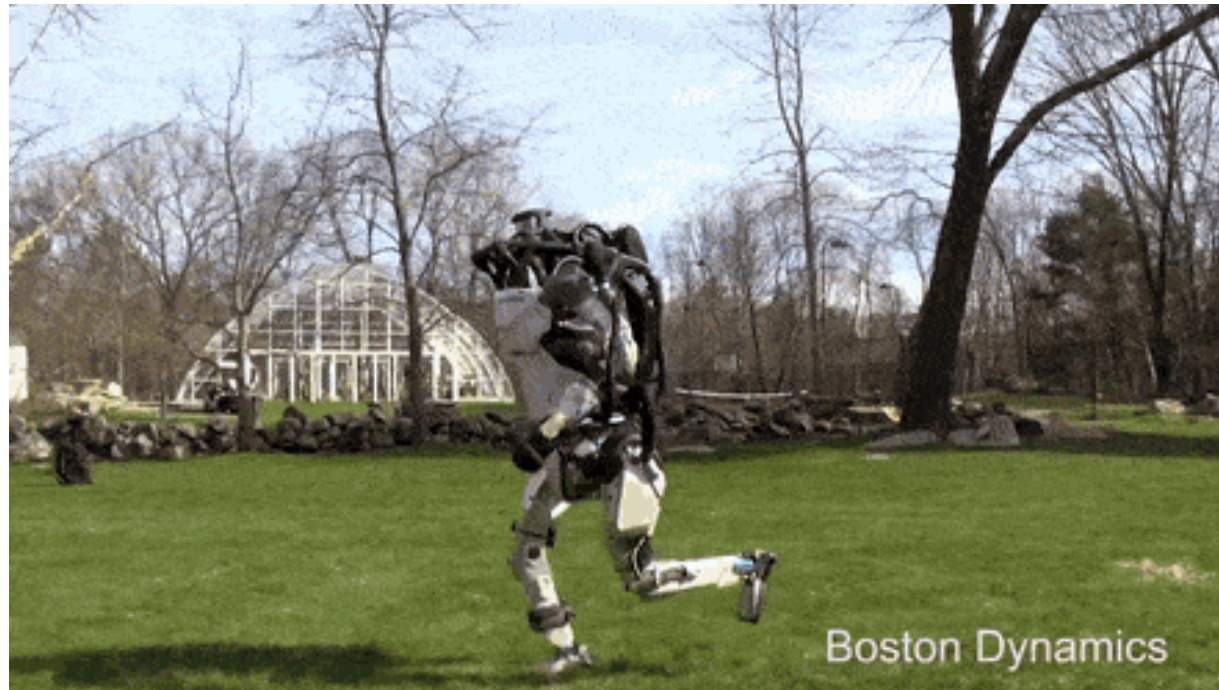
Recurrent neural networks in practice



Q: what is the action?

Runn

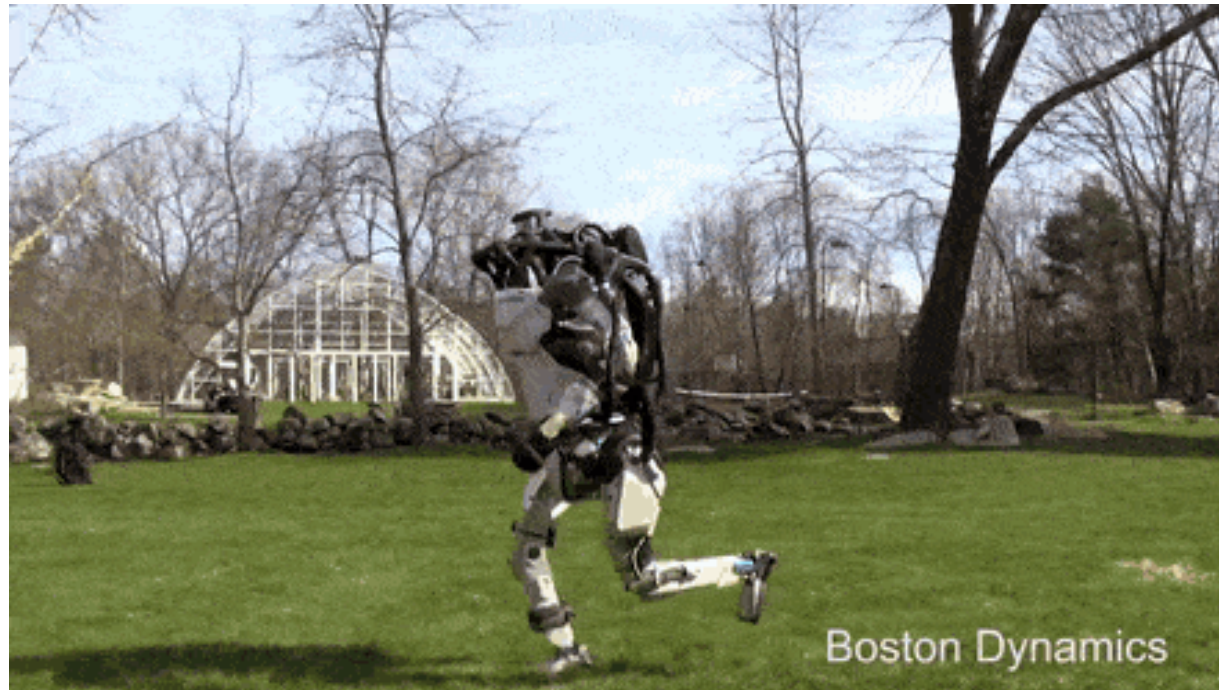
Recurrent neural networks in practice



Q: what is the action?

Runni

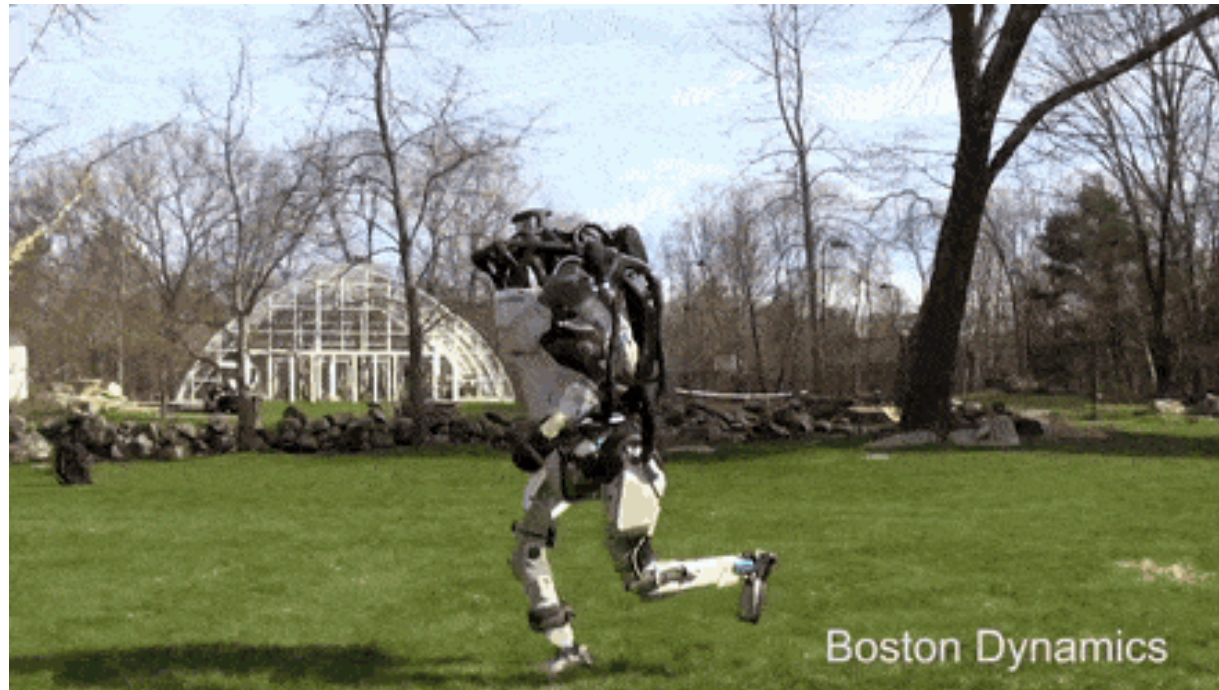
Recurrent neural networks in practice



Q: what is the action?

Runnin

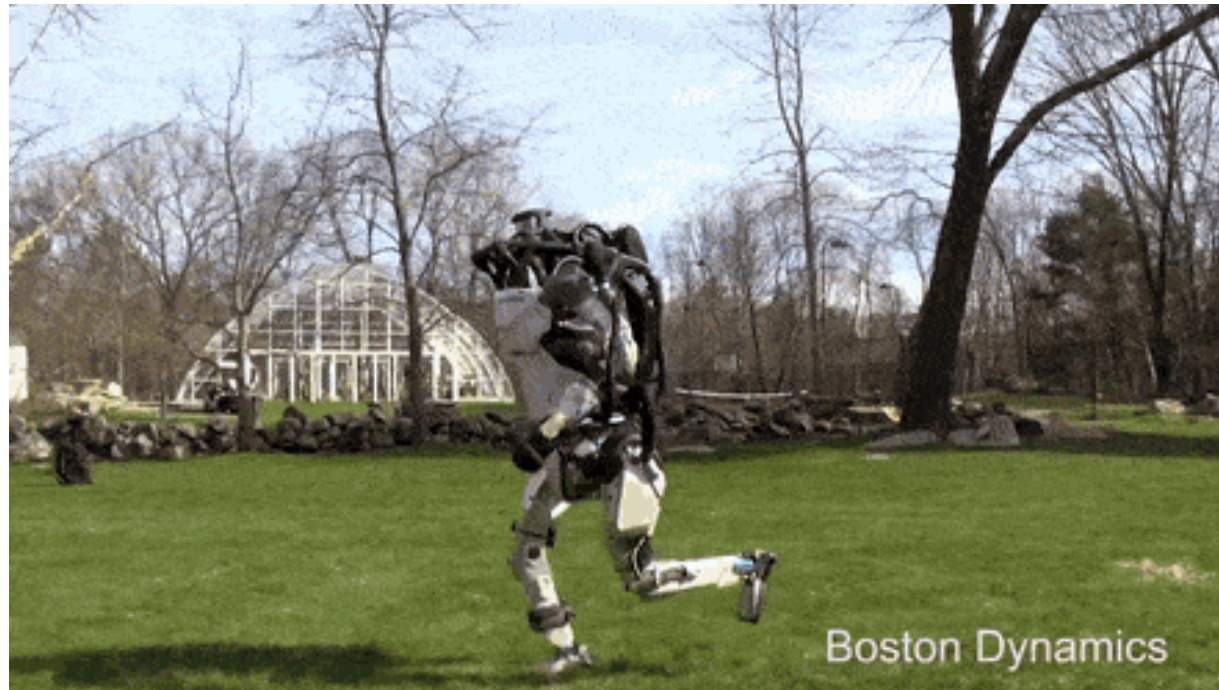
Recurrent neural networks in practice



Q: what is the action?

Running

Recurrent neural networks in practice



Q: what is the action?

Running ← Sequence data

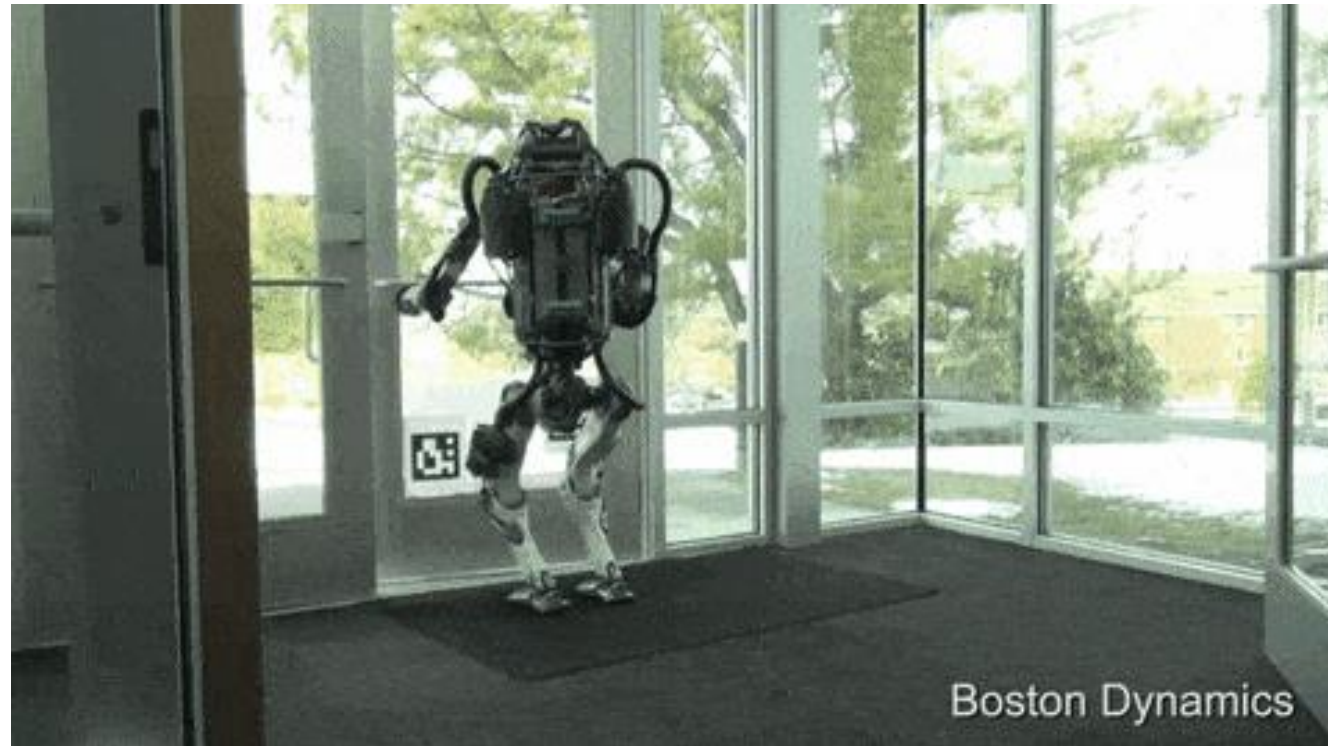
Recurrent neural networks in practice



Q: what is the action?

Opening a door

Recurrent neural networks in practice



Video captioning:
Generate captions

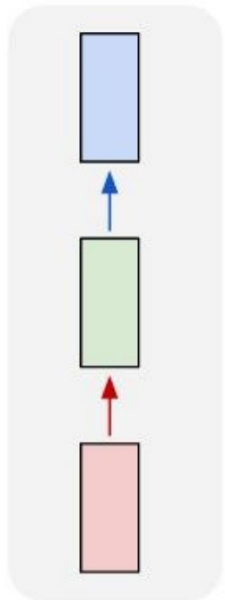
Q: what is the action?

Opening a door

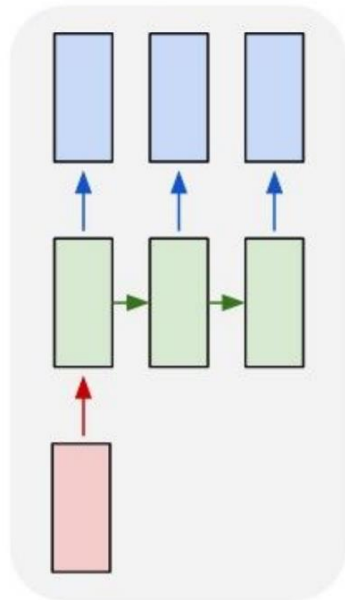
Recurrent neural networks

What real applications?

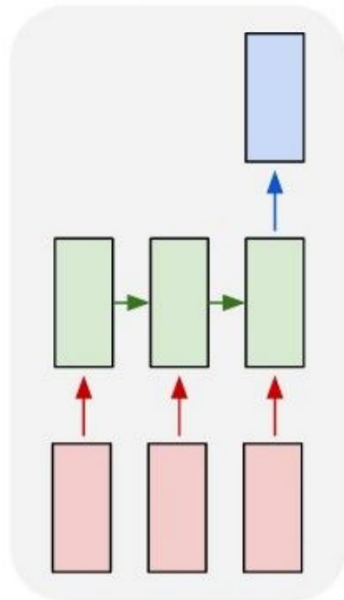
one to one



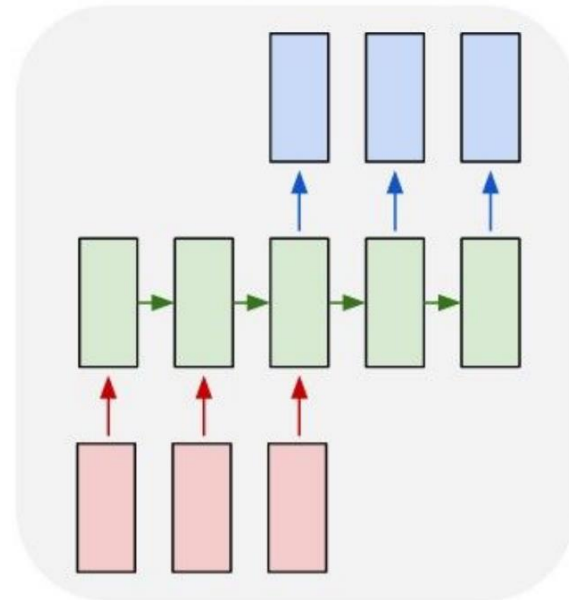
one to many



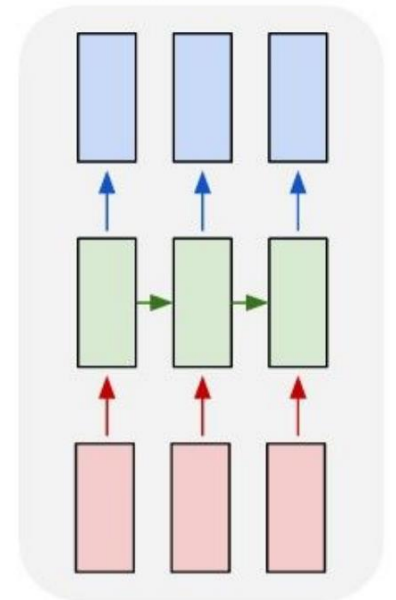
many to one



many to many



many to many



Recurrent neural networks

What real applications?

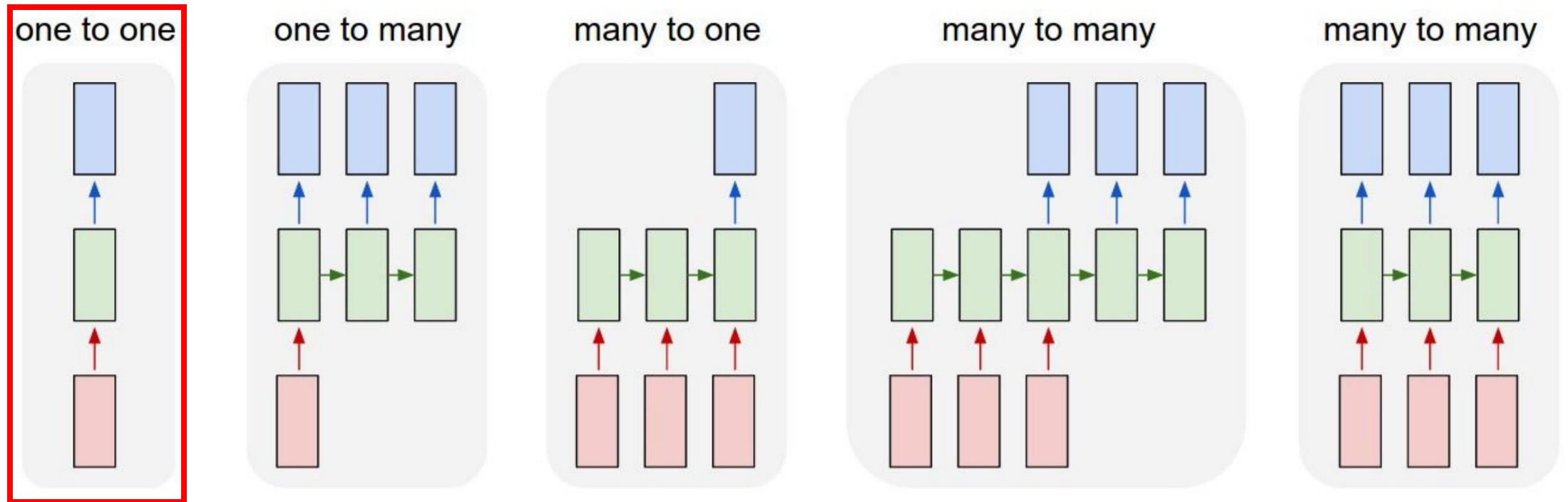
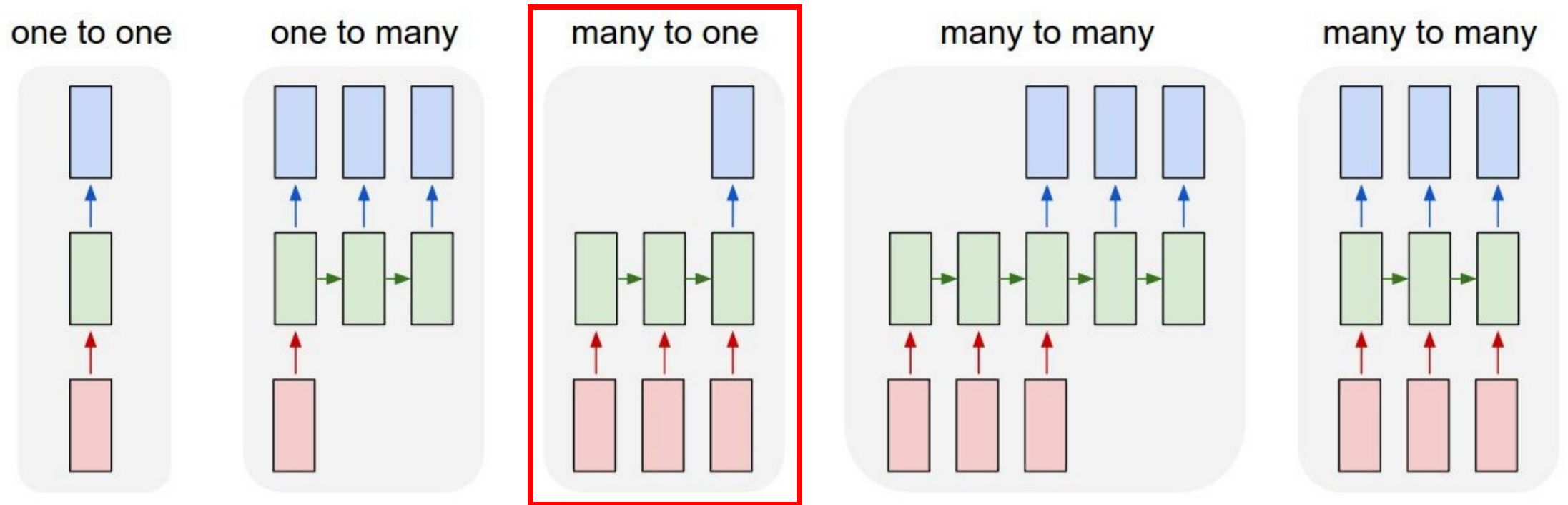


Image classification

Recurrent neural networks

What real applications?

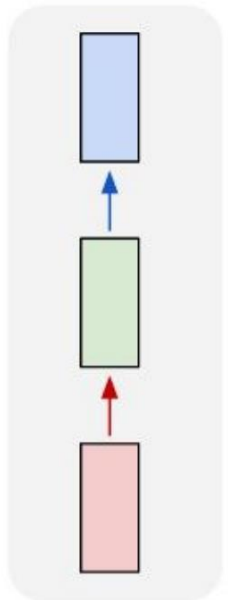


Action recognition

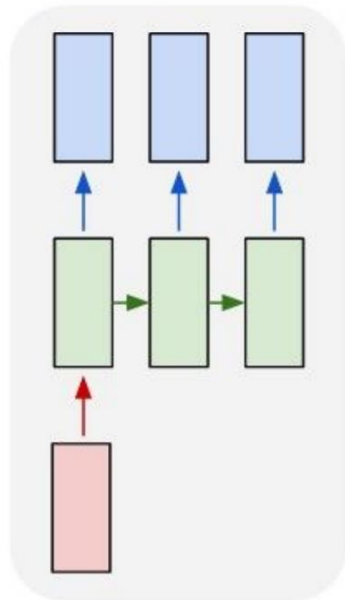
Recurrent neural networks

What real applications?

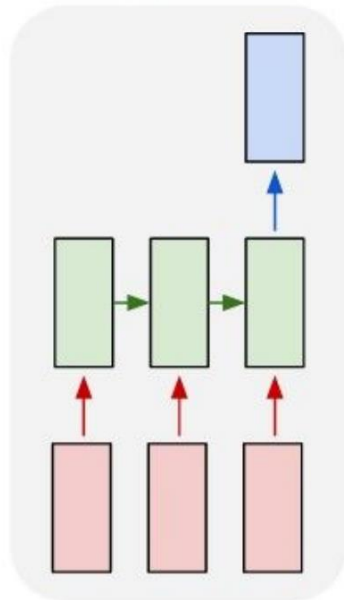
one to one



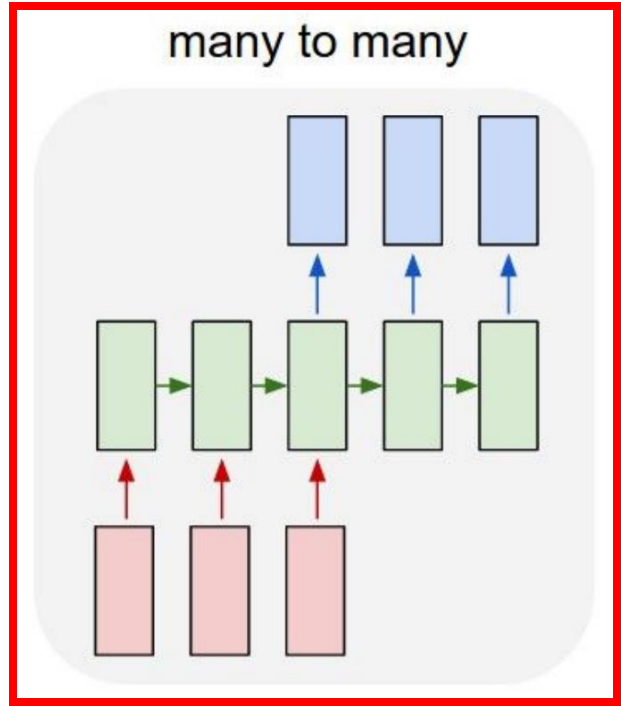
one to many



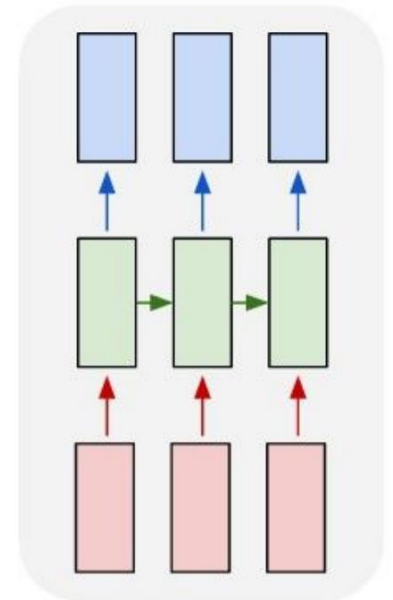
many to one



many to many



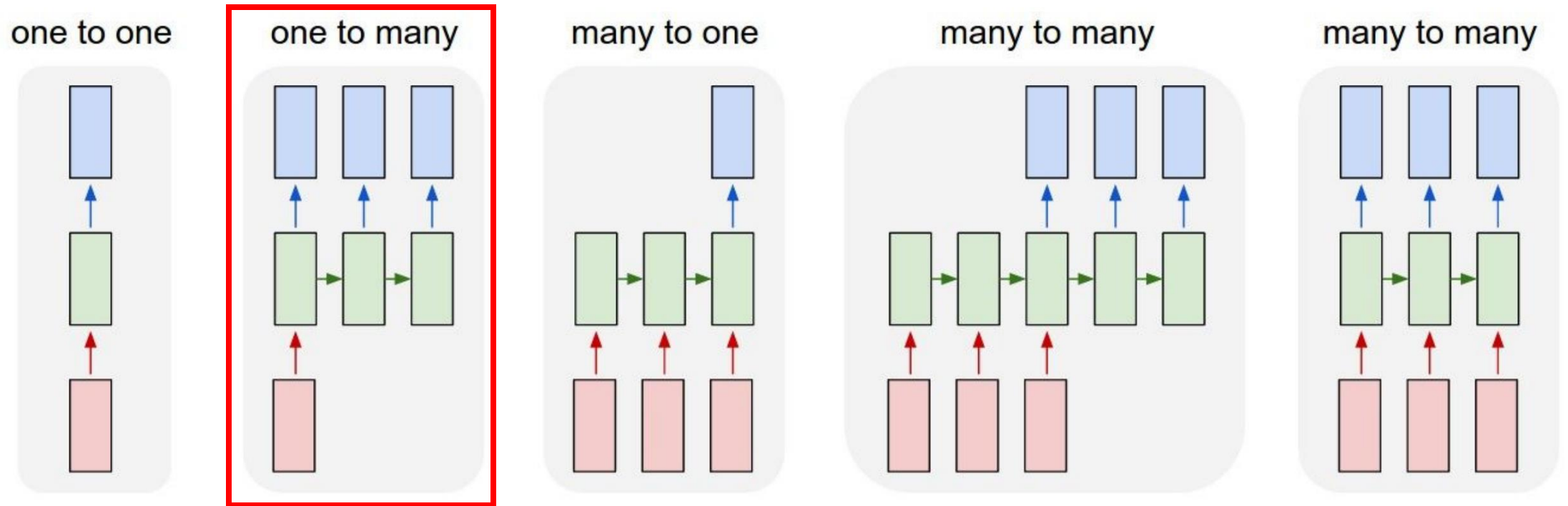
many to many



Video captioning

Recurrent neural networks

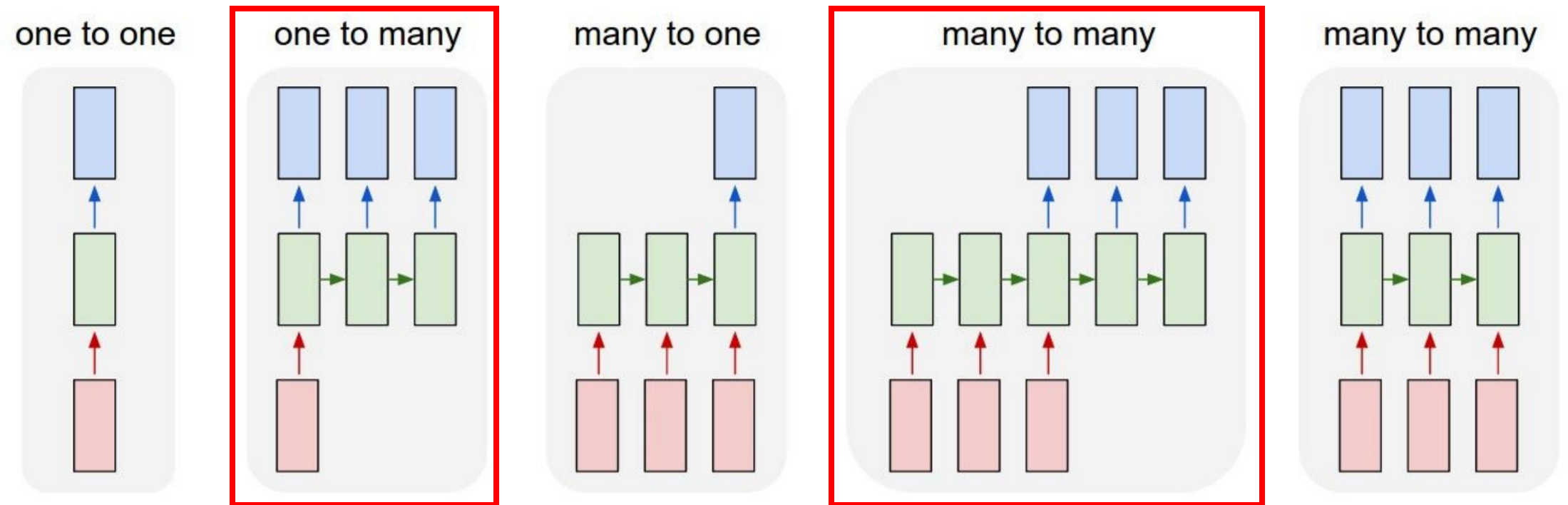
What real applications?



Q: what application?

Recurrent neural networks

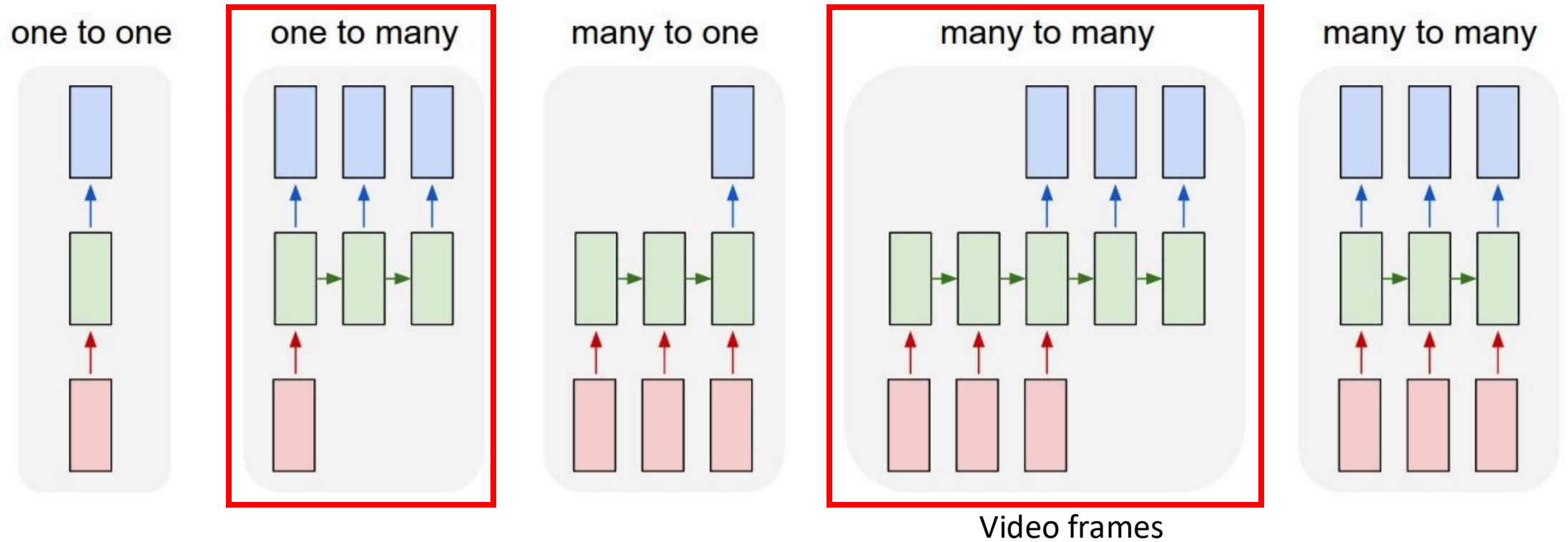
What real applications?



Q: what application?

Recurrent neural networks

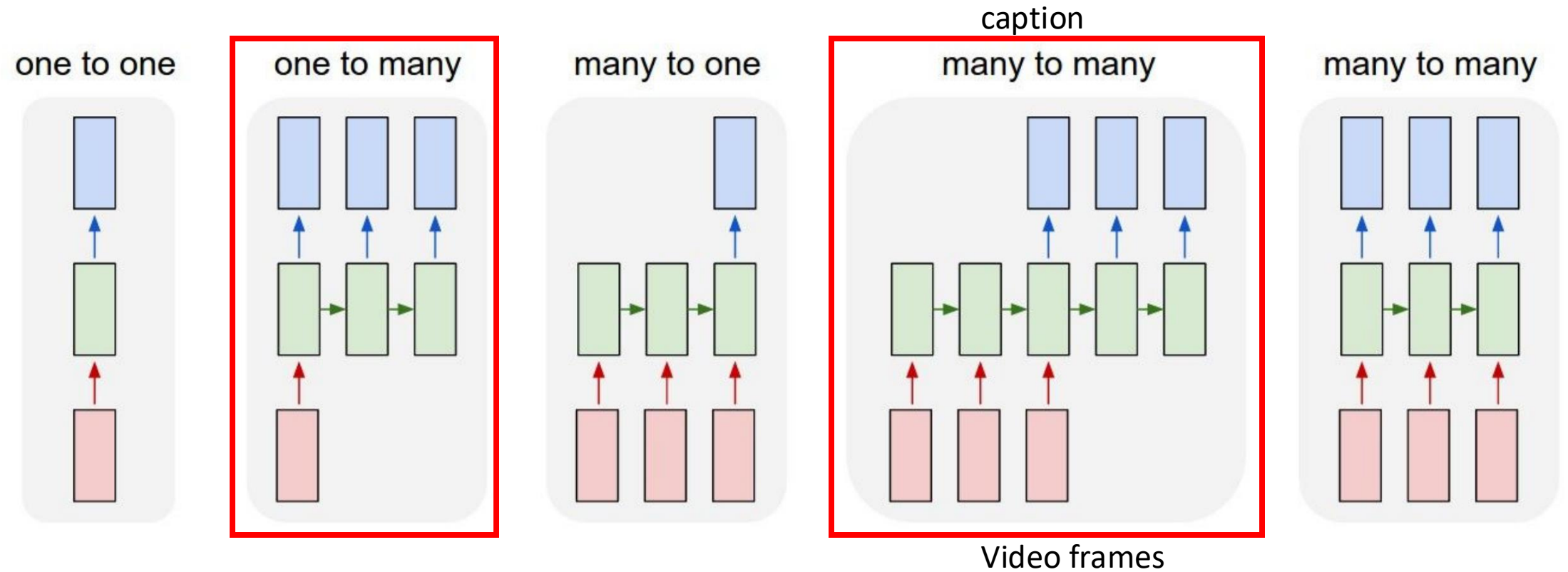
What real applications?



Q: what application?

Recurrent neural networks

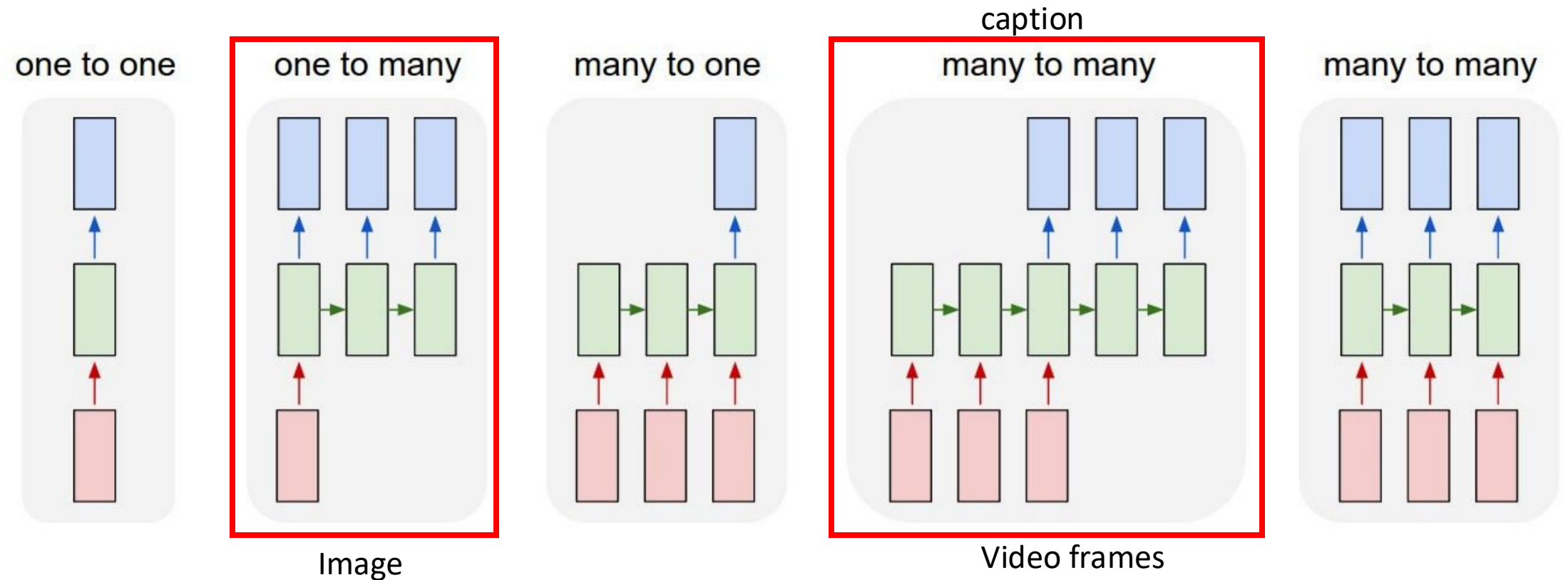
What real applications?



Q: what application?

Recurrent neural networks

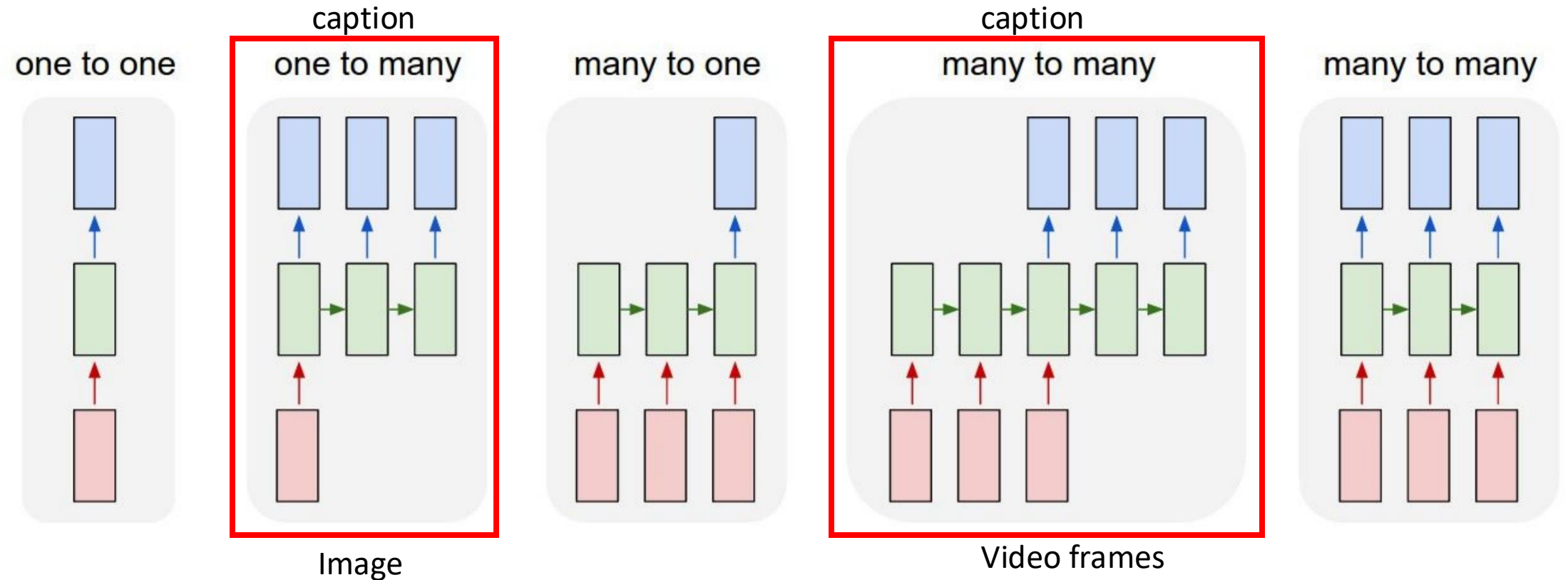
What real applications?



Q: what application?

Recurrent neural networks

What real applications?



Q: what application?

Image captioning

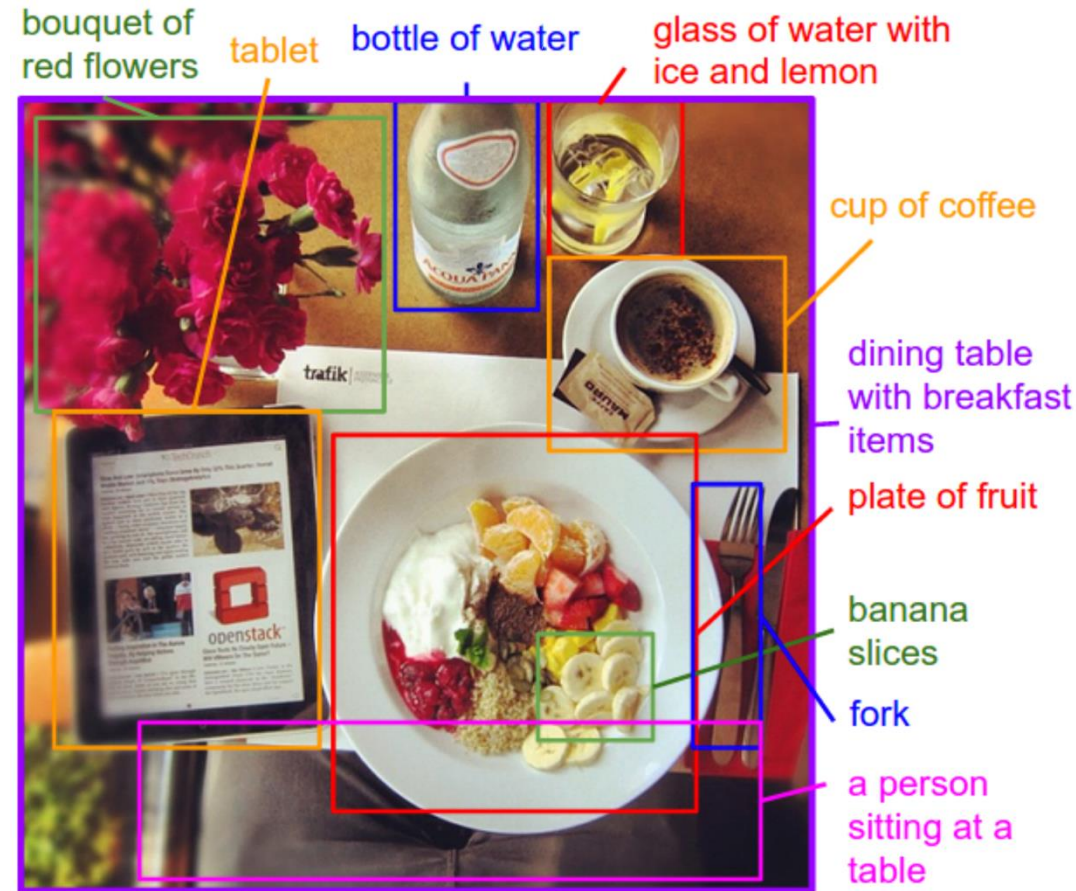


Figure from Karpathy, Andrej, and Li Fei-Fei. "Deep visual-semantic alignments for generating image descriptions." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3128-3137. 2015.

Image captioning

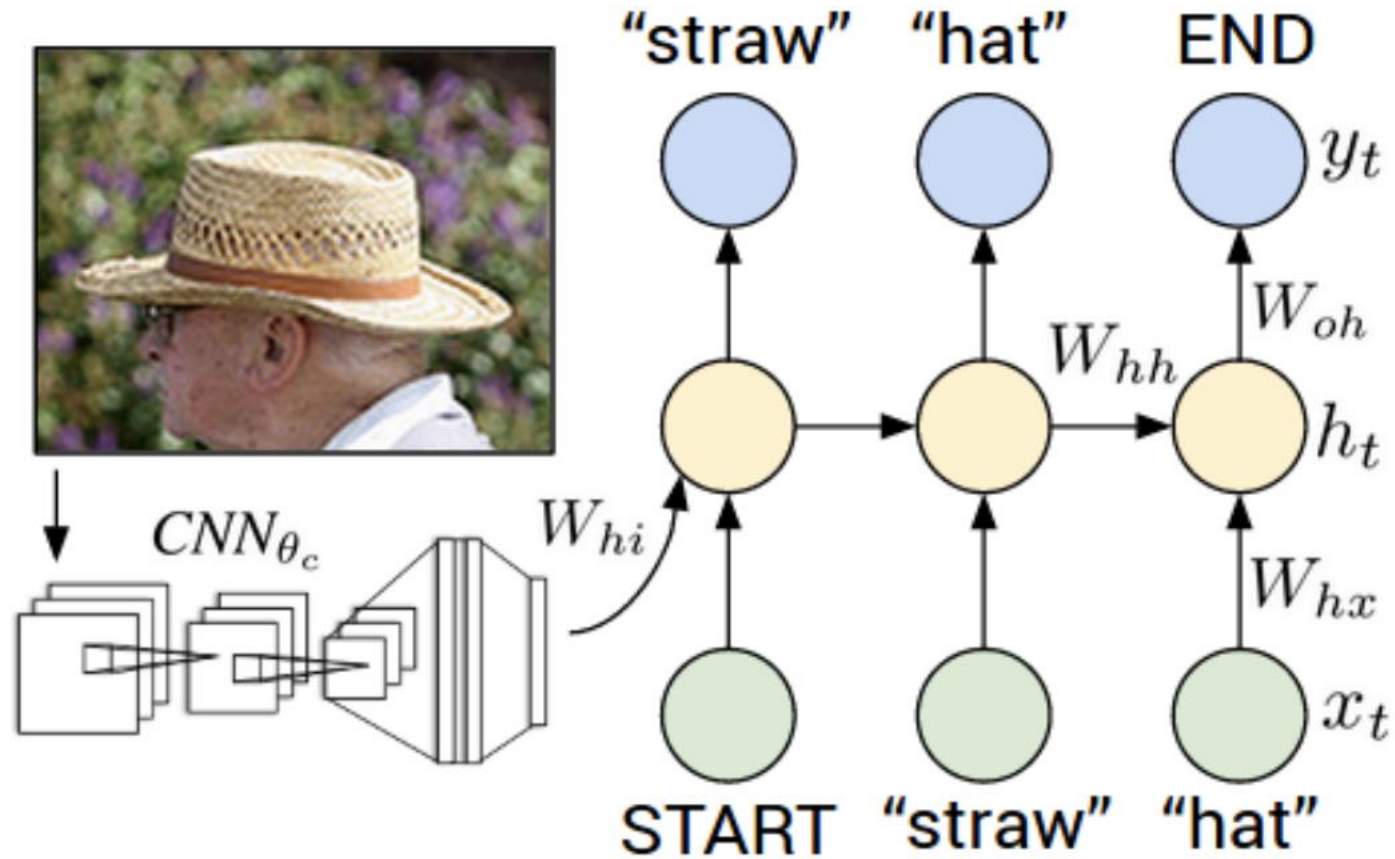
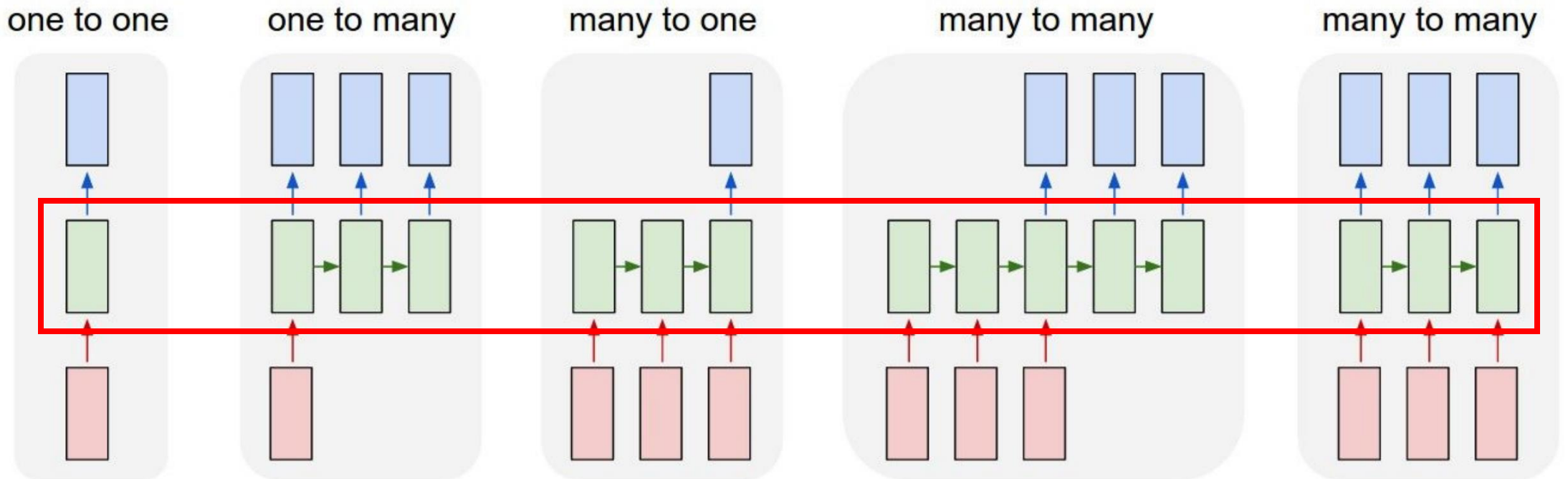


Figure from Karpathy, Andrej, and Li Fei-Fei. "Deep visual-semantic alignments for generating image descriptions." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3128-3137. 2015.

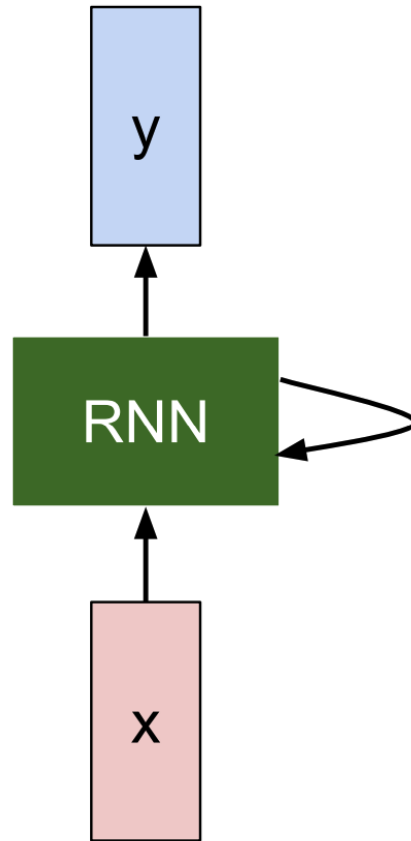
Recurrent neural networks

What's the key?

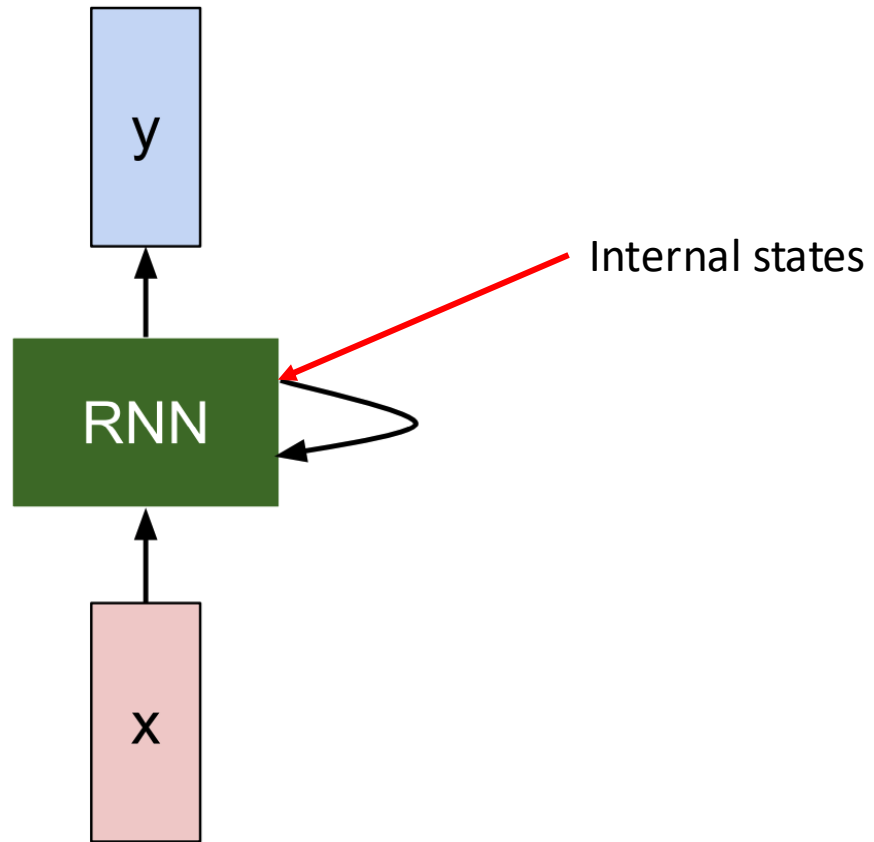


Q: what application?

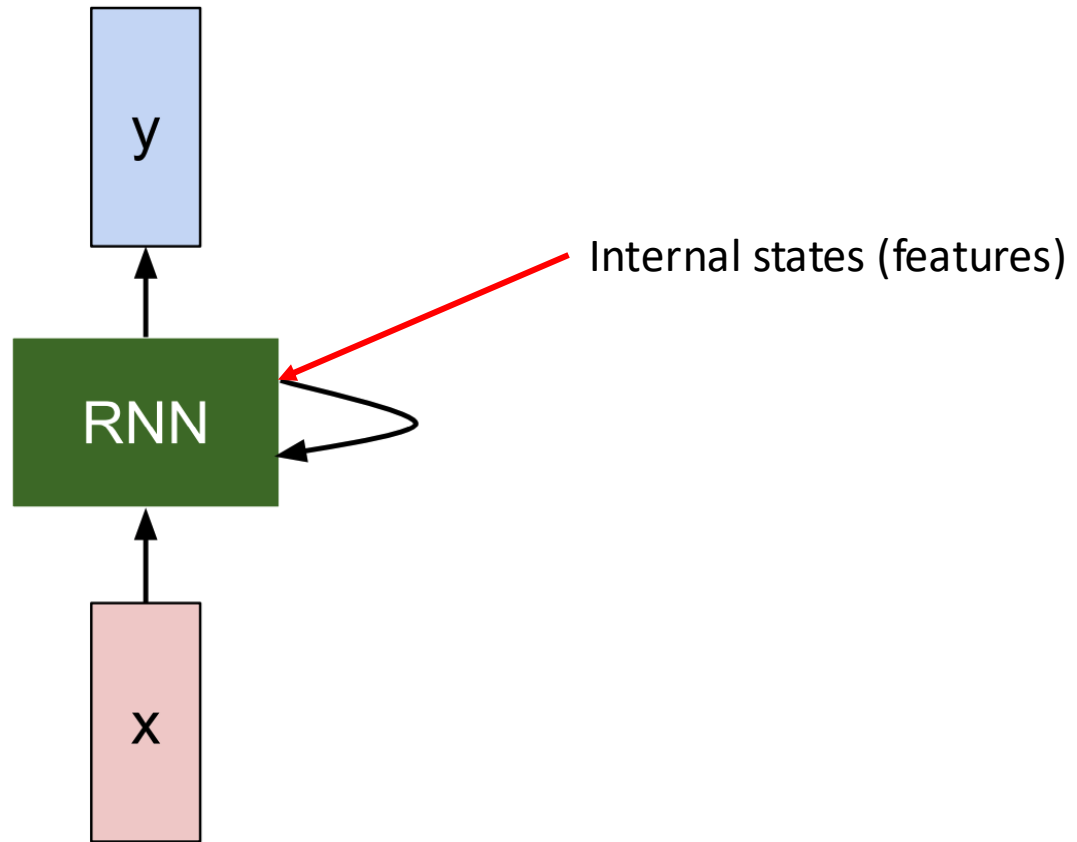
Recurrent neural networks



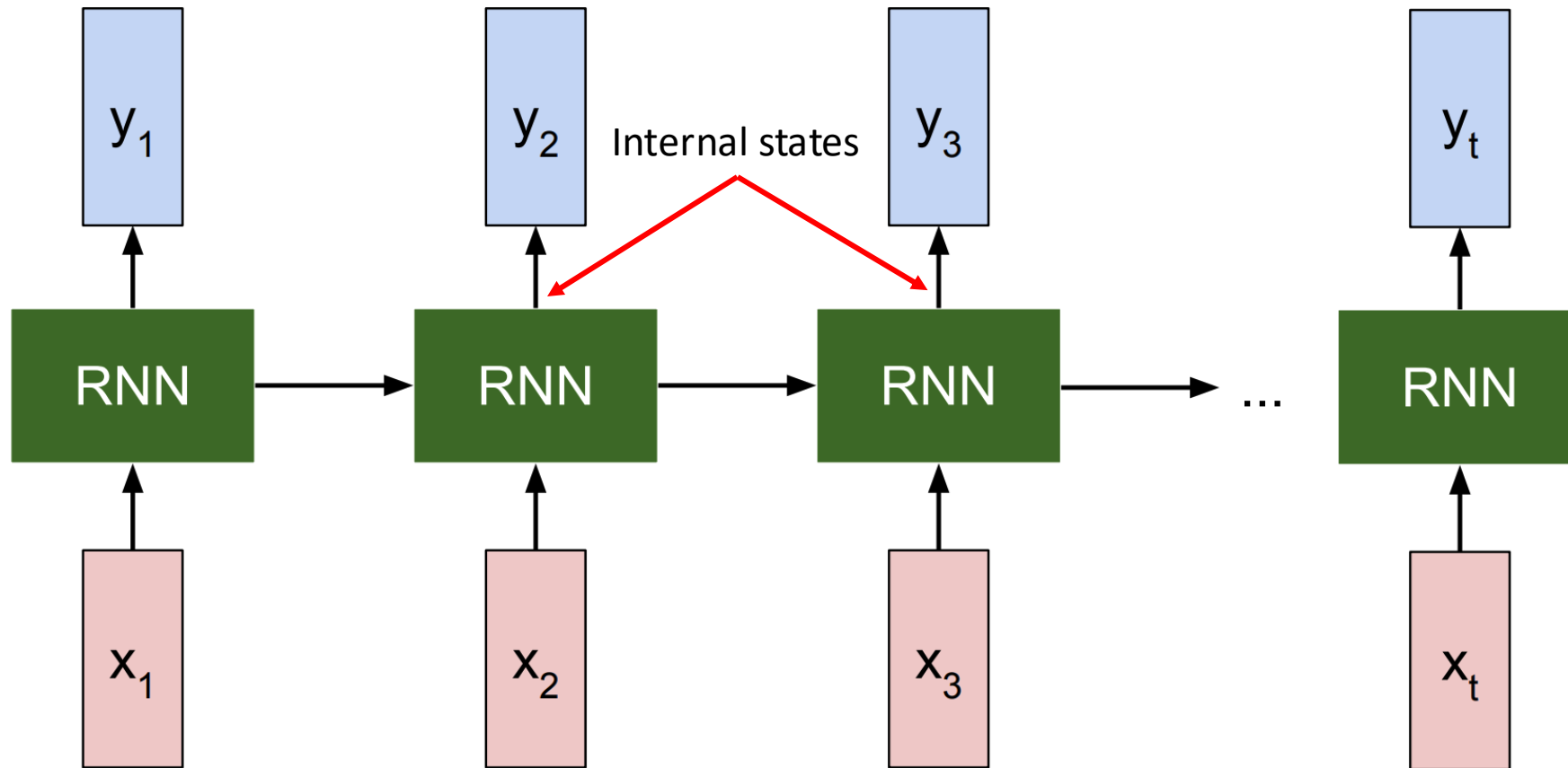
Recurrent neural networks



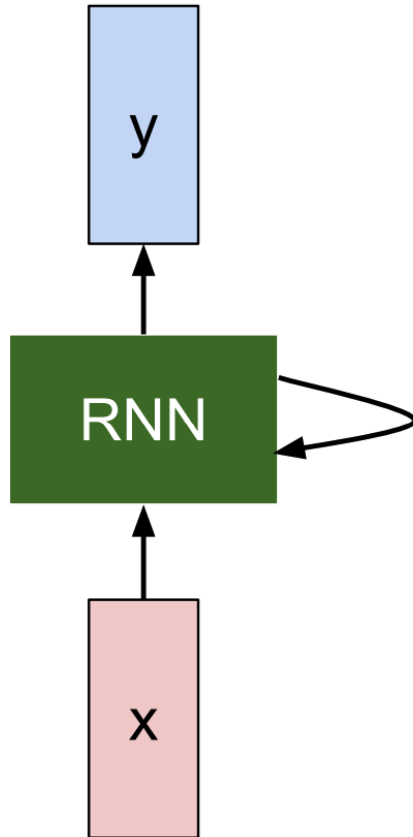
Recurrent neural networks



Recurrent neural networks



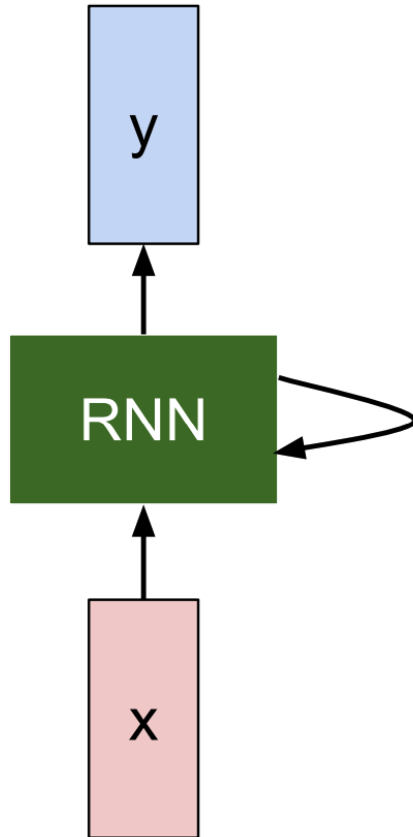
Recurrent neural networks



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

new state some function with parameters W old state input vector at some time step

Recurrent neural networks

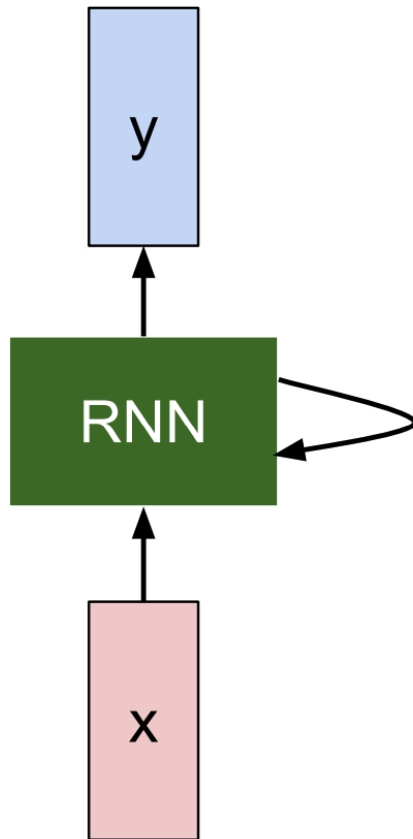


$$h_t = f_W(h_{t-1}, x_t)$$

new state some function with parameters W old state input vector at some time step

$\begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix}$

Recurrent neural networks



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

new state

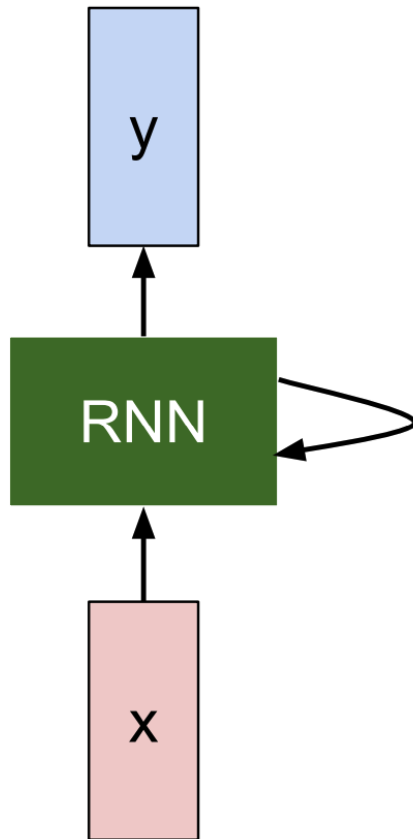
some function with parameters W

old state

input vector at some time step

$$W * \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix}$$

Recurrent neural networks

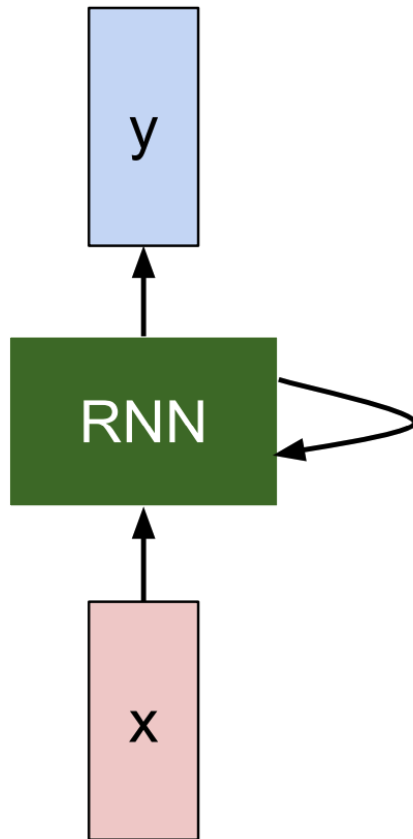


$$h_t = f \left[W * \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} \right]$$

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

new state / some function with parameters W / old state / input vector at some time step

Recurrent neural networks



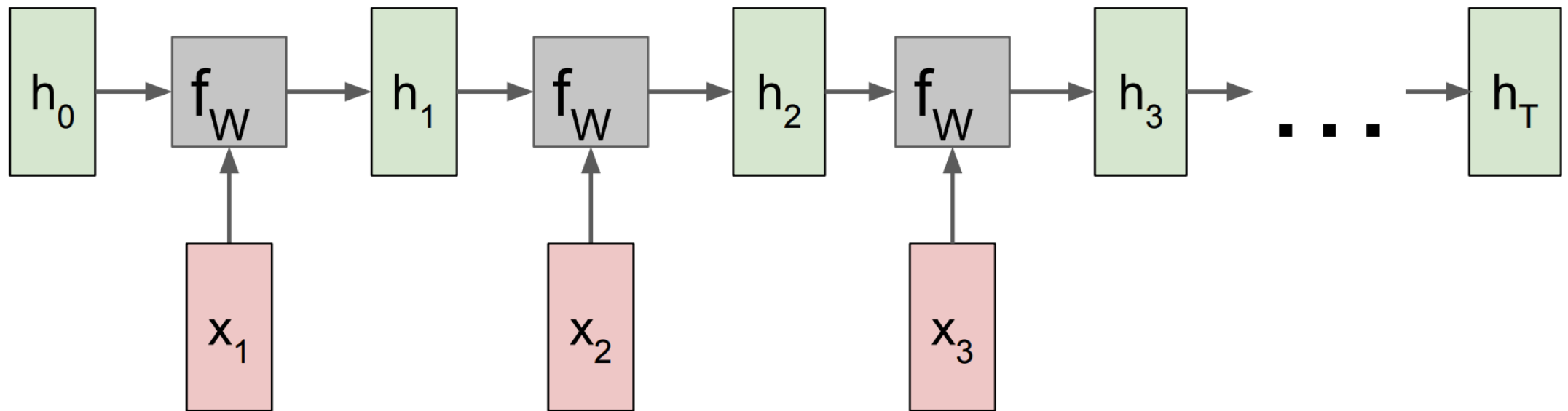
$$h_t = f \left[W * \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} \right]$$

$$f = \tanh(\cdot)$$

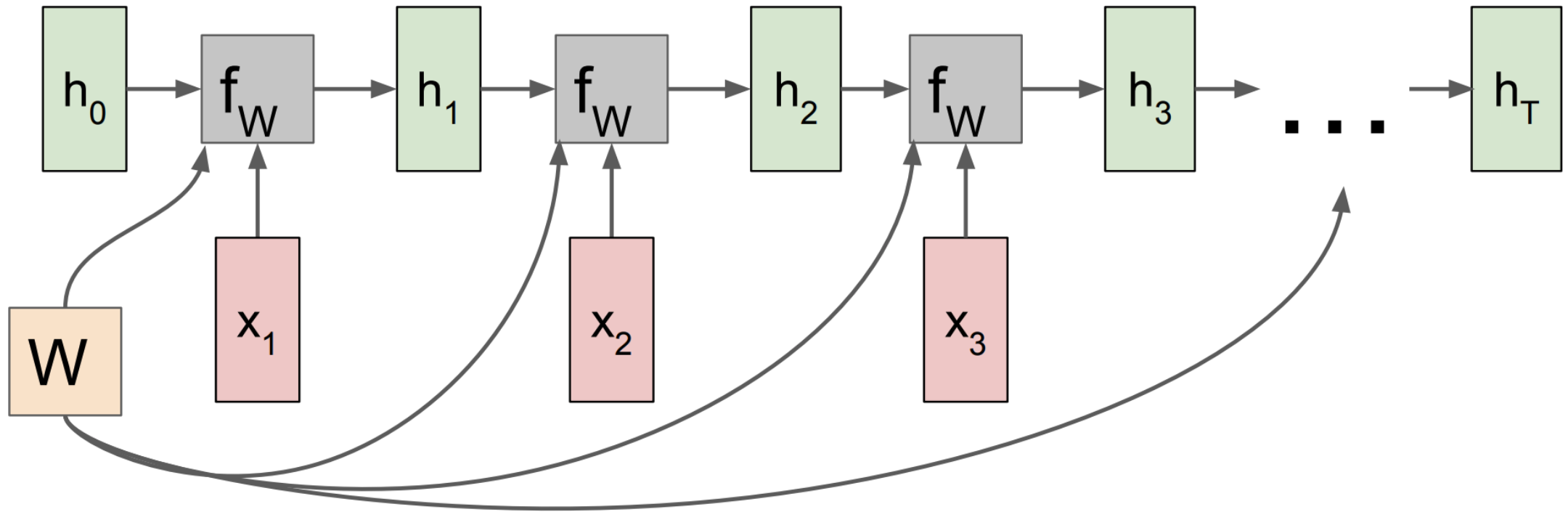
$$\boxed{h_t} = \boxed{f_W}(\boxed{h_{t-1}}, \boxed{x_t})$$

new state / some function with parameters W / old state / input vector at some time step

Recurrent neural networks

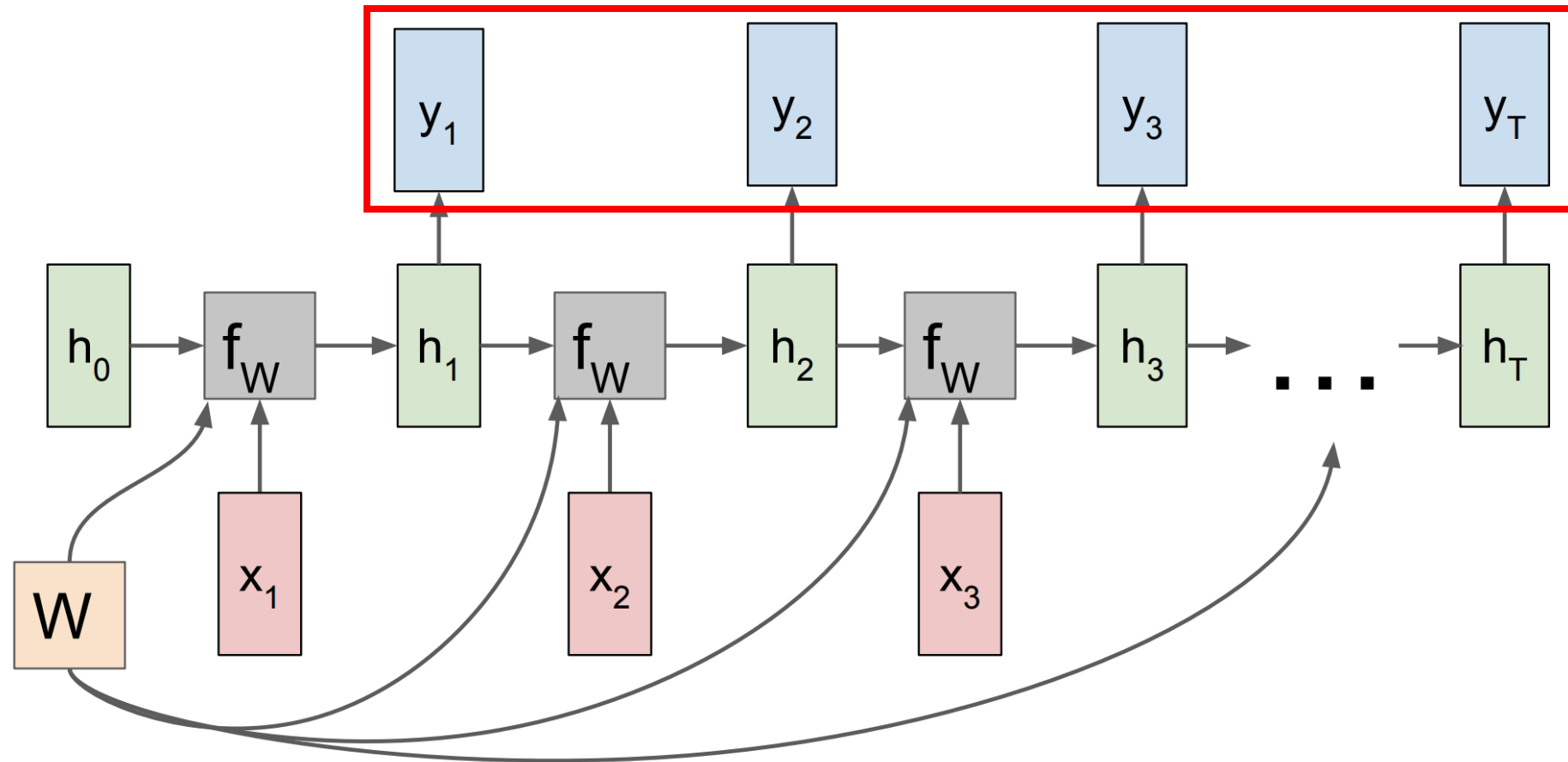


Recurrent neural networks



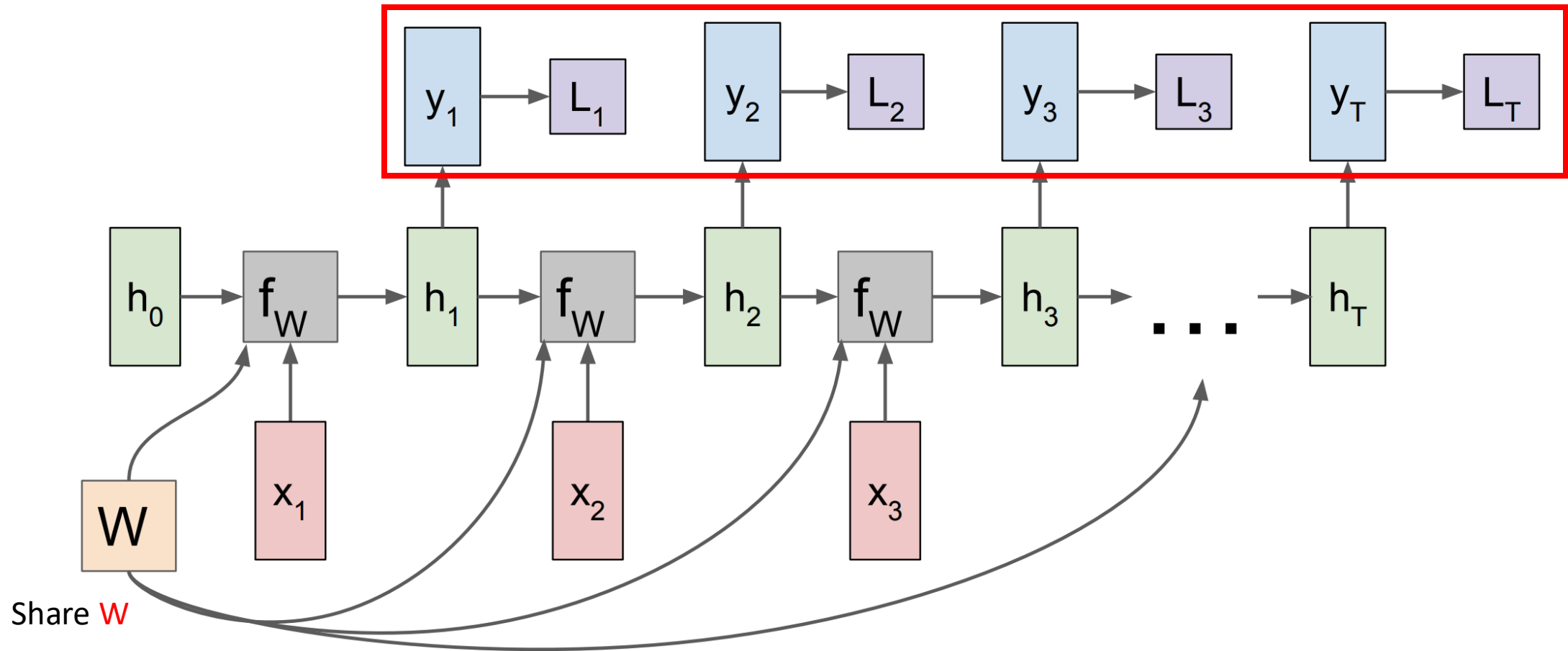
Share W

Recurrent neural networks

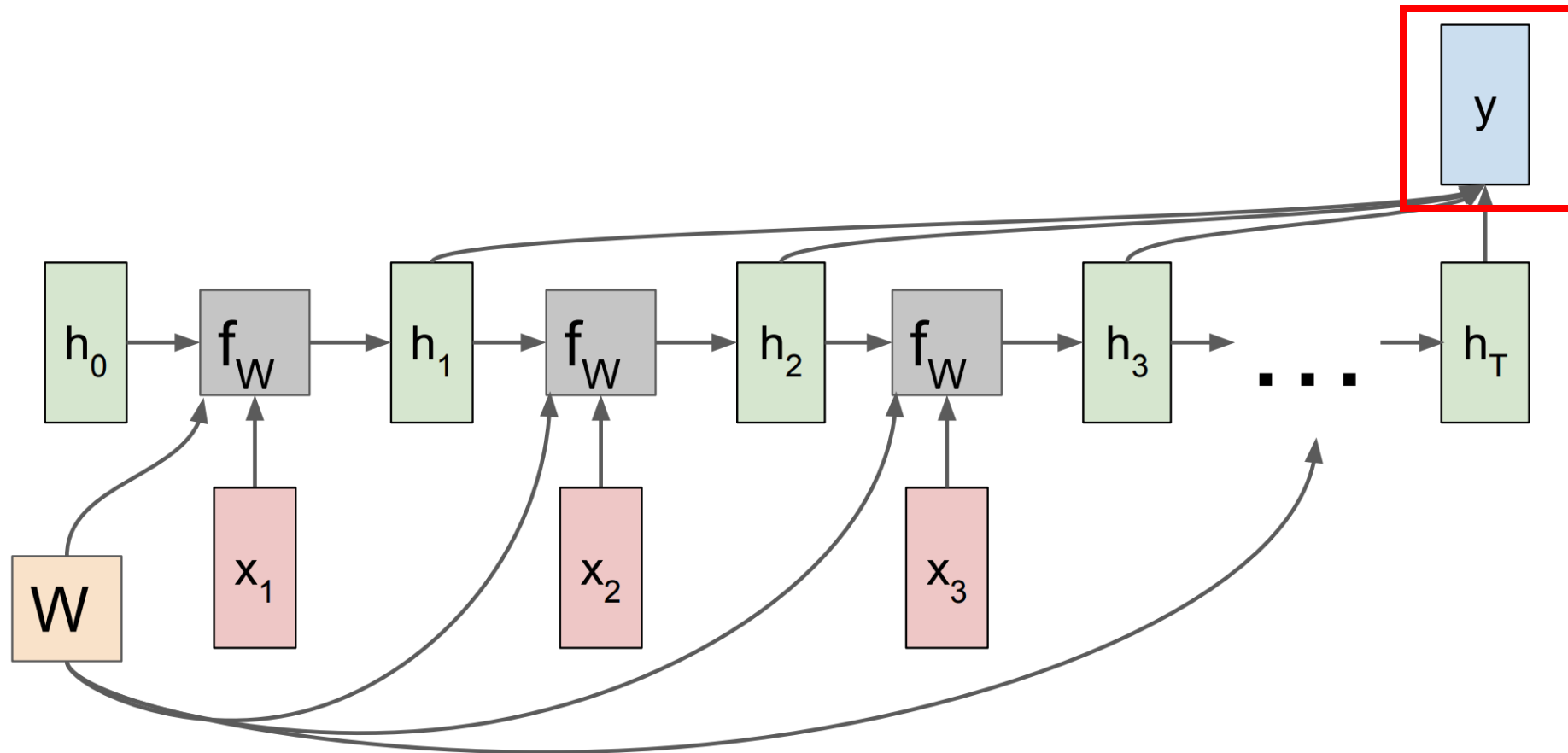


Share W

Recurrent neural networks

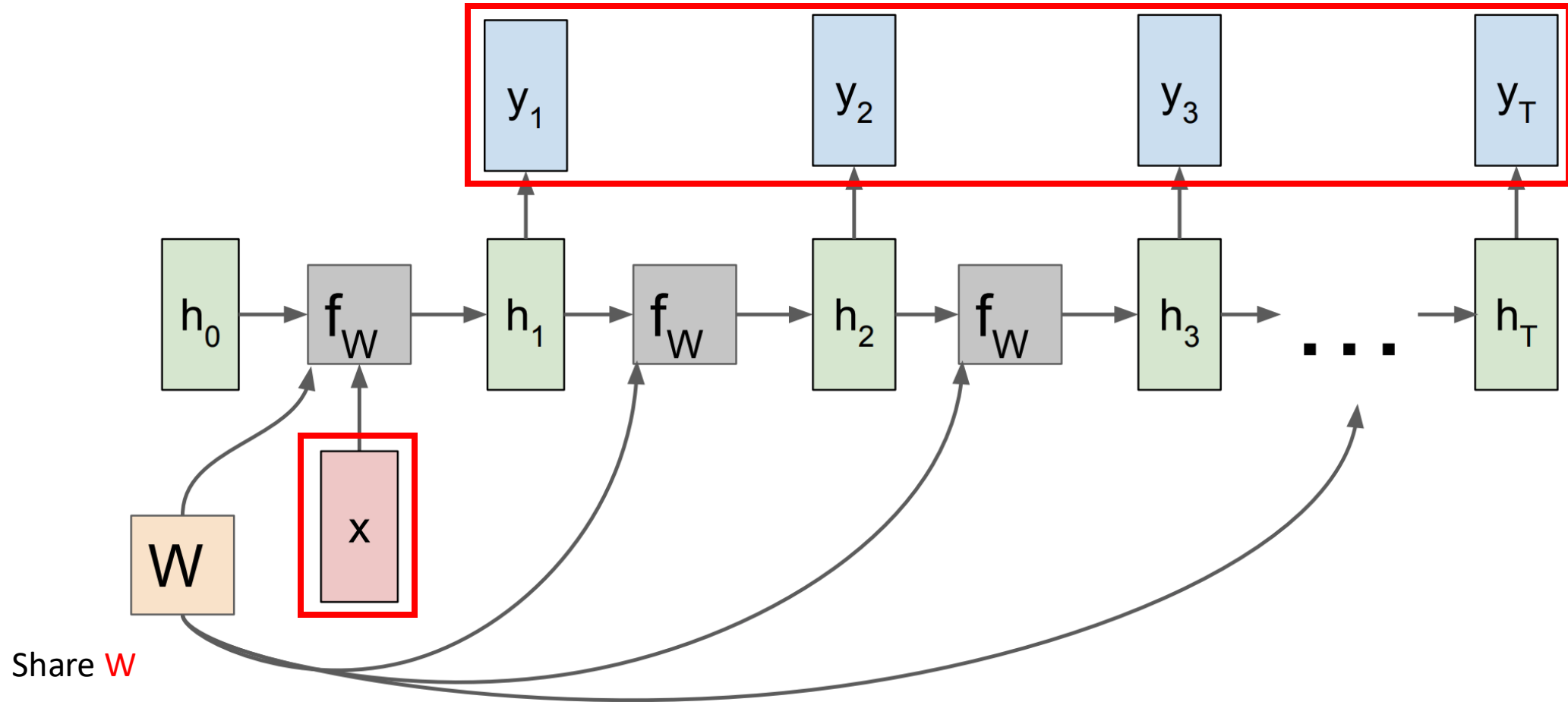


Recurrent neural networks

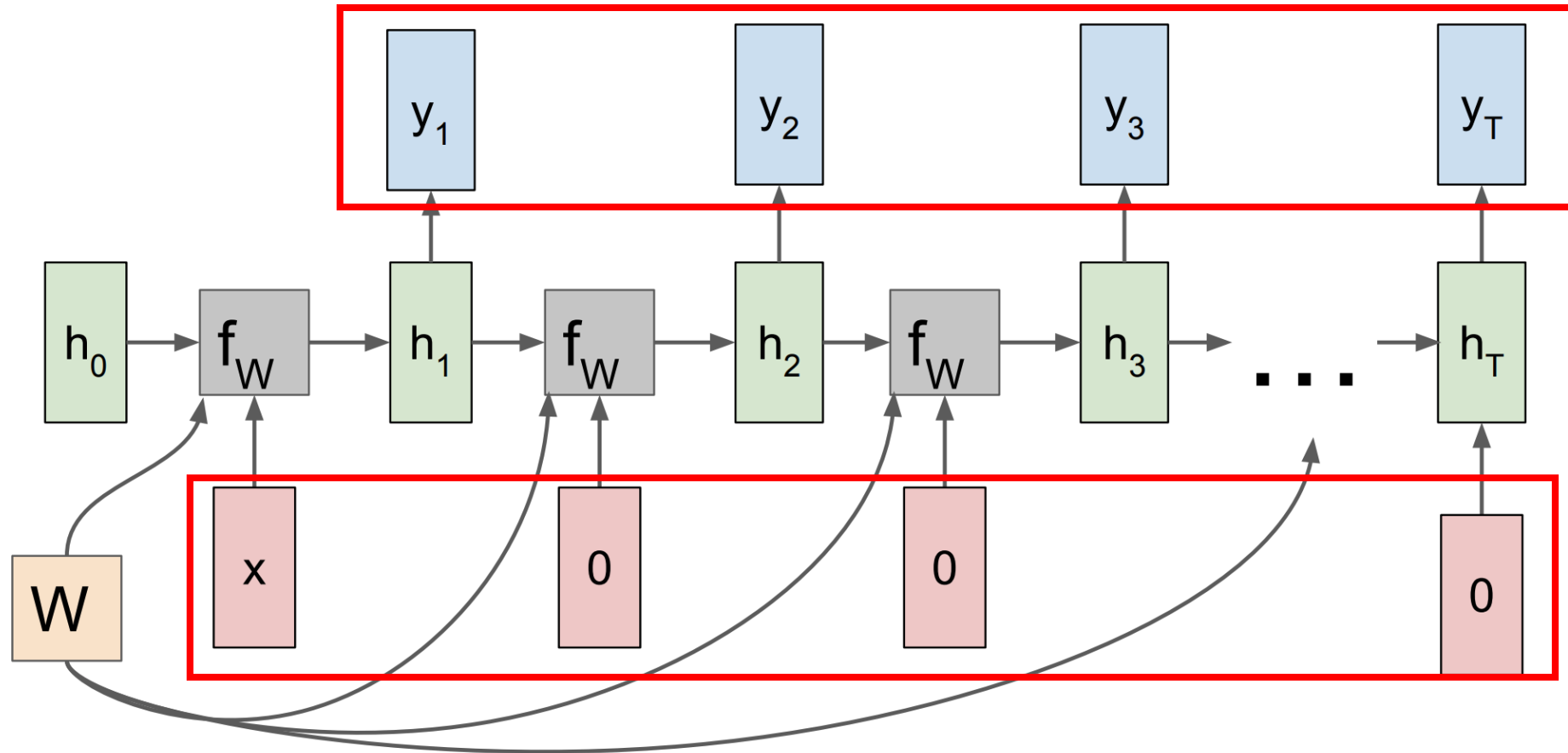


Share W

Recurrent neural networks



Recurrent neural networks



Share W

RNNs for language model

- Guess the word:
 - h

RNNs for language model

- Guess the word:
 - he

RNNs for language model

- Guess the word:
 - hel

RNNs for language model

- Guess the word:
 - hell

RNNs for language model

- Guess the word:
 - hello

RNNs for language model

- Guess the word:
 - hello
 - net

RNNs for language model

- Guess the word:
 - hello
 - netw

RNNs for language model

- Guess the word:
 - hello
 - netwo

RNNs for language model

- Guess the word:
 - hello
 - network

RNNs for language model

- Guess the word:
 - hello
 - network
 - I

RNNs for language model

- Guess the word:
 - hello
 - network
 - lan

RNNs for language model

- Guess the word:
 - hello
 - network
 - langu

RNNs for language model

- Guess the word:
 - hello
 - network
 - languag

RNNs for language model

- Guess the word:
 - hello
 - network
 - language

RNNs for language model

- Guess the word:
 - hello
 - network
 - language
- Sequence data: predict the next value

RNNs for language model

- Guess the word:
 - hello
 - network
 - language
- Sequence data: predict the next value
 - n
 - n

RNNs for language model

- Guess the word:
 - hello
 - network
 - language
- Sequence data: predict the next value
 - ne
 - ne

RNNs for language model

- Guess the word:
 - hello
 - network
 - language
- Sequence data: predict the next value
 - neu
 - net

RNNs for language model

- Guess the word:
 - hello
 - network
 - language
- Sequence data: predict the next value
 - neur
 - netw

RNNs for language model

- Guess the word:
 - hello
 - network
 - language
- Sequence data: predict the next value
 - neura
 - netwo

RNNs for language model

- Guess the word:
 - hello
 - network
 - language
- Sequence data: predict the next value
 - neural
 - network

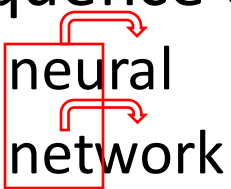
RNNs for language model

- Guess the word:
 - hello
 - network
 - language
- Sequence data: predict the next value
 - neural
 - network

RNNs for language model

- Guess the word:
 - hello
 - network
 - language
- Sequence data: predict the next value
 - neural
 - network

RNNs for language model

- Guess the word:
 - hello
 - network
 - language
- Sequence data: predict the next value
 -  neural Information flow
 - network

Character-level language model

- Vocabulary: {a, b, ..., z}

Character-level language model

- Vocabulary: {a, b, ..., z}
- Given a sequence of character:

Character-level language model

- Vocabulary: {a, b, ..., z}
- Given a sequence of character:
 - hellx
 - mornixx
 - languaxx
 - neurxx
 - netwxxx
 - ...

Character-level language model

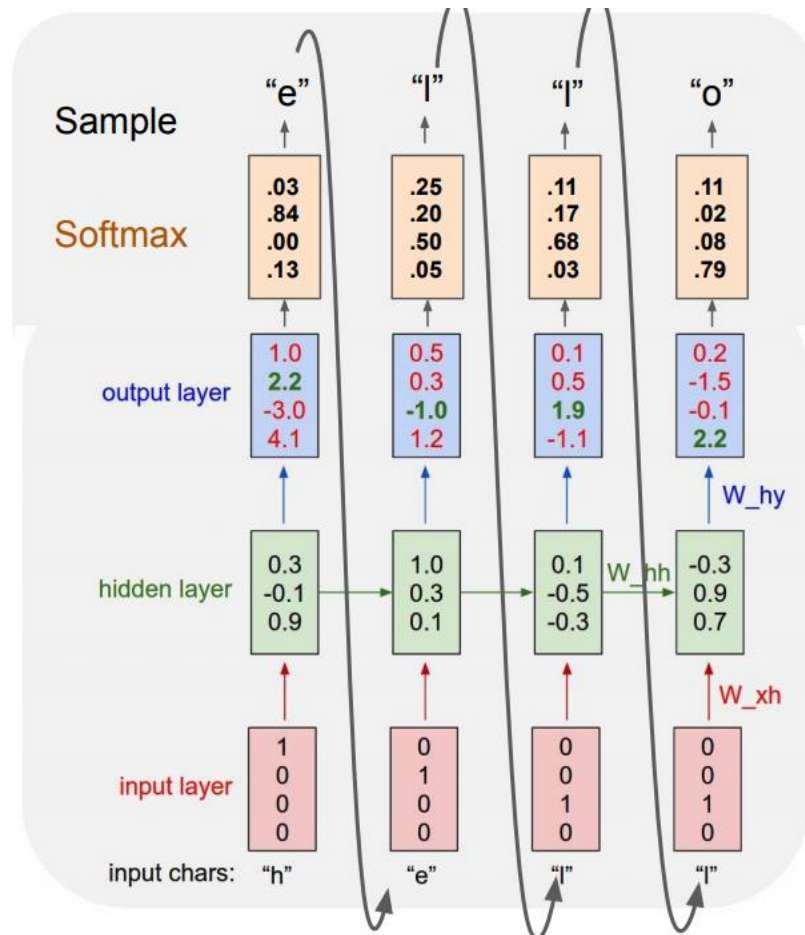
- Vocabulary: {a, b, ..., z}
- Given a sequence of character:
 - hellx → hello
 - mornixx → morning
 - languaxx → language
 - neurxx → neural
 - netwxxx → network
 - ...

Character-level language model

- Vocabulary: {h, e, l, o}

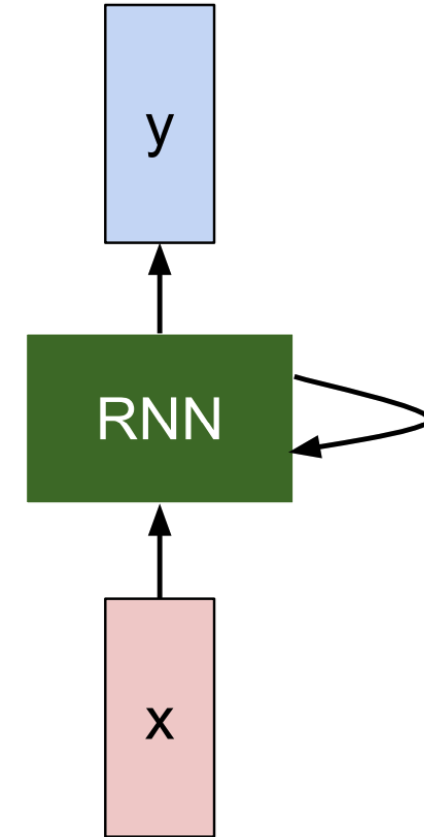
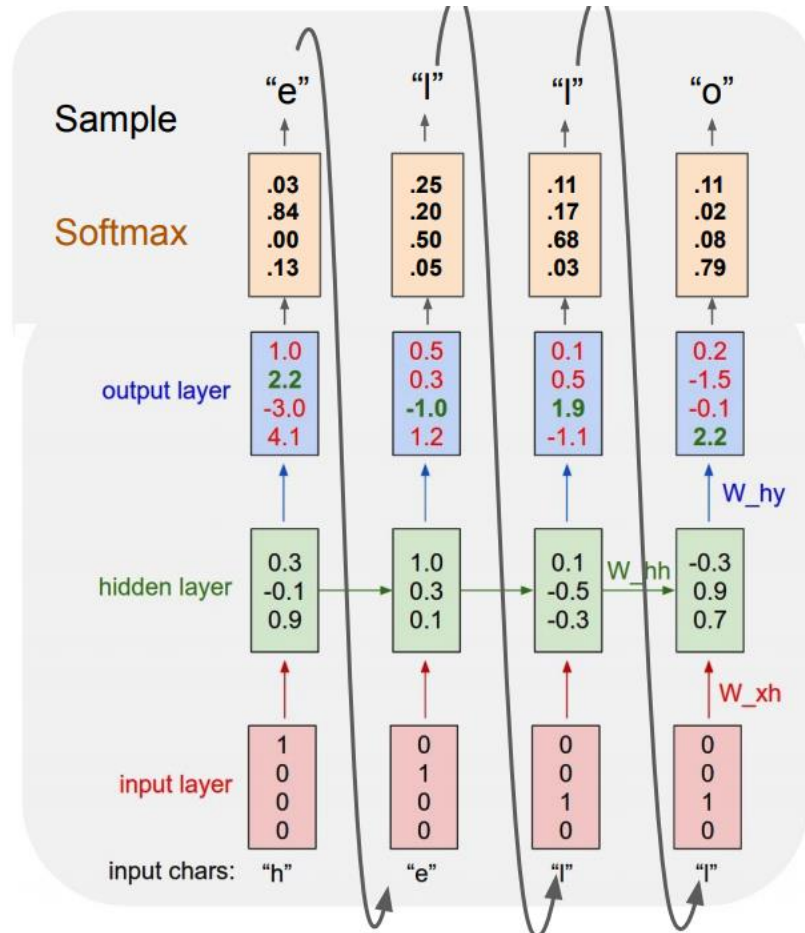
Character-level language model

- Vocabulary: {h, e, l, o}



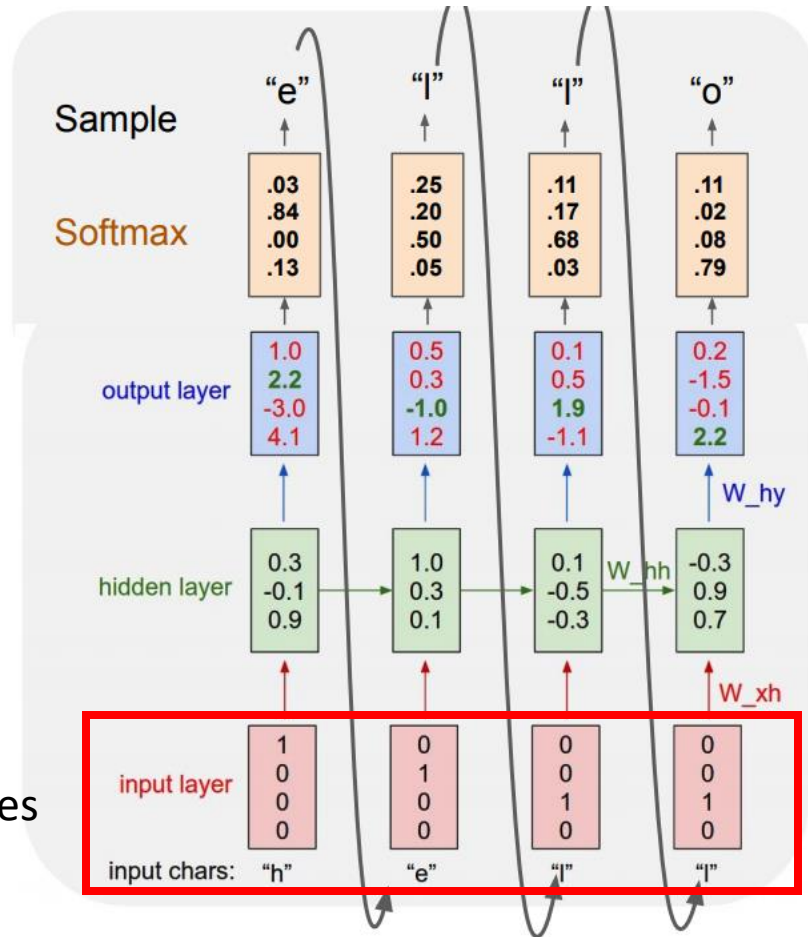
Character-level language model

- Vocabulary: {h, e, l, o}

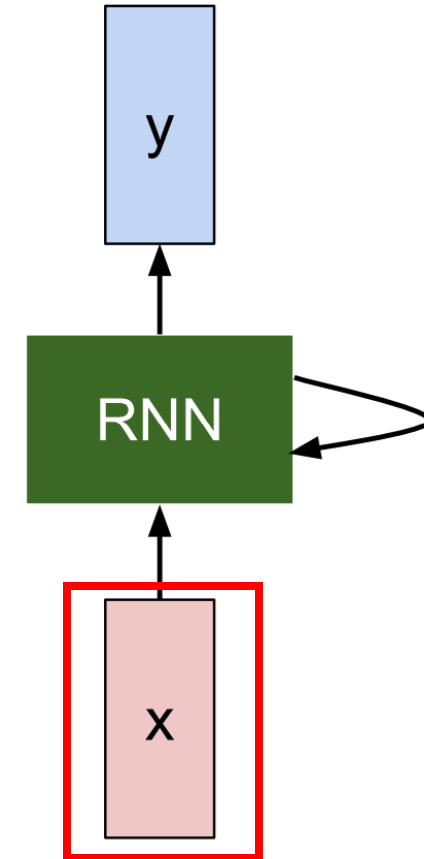


Character-level language model

- Vocabulary: {h, e, l, o}

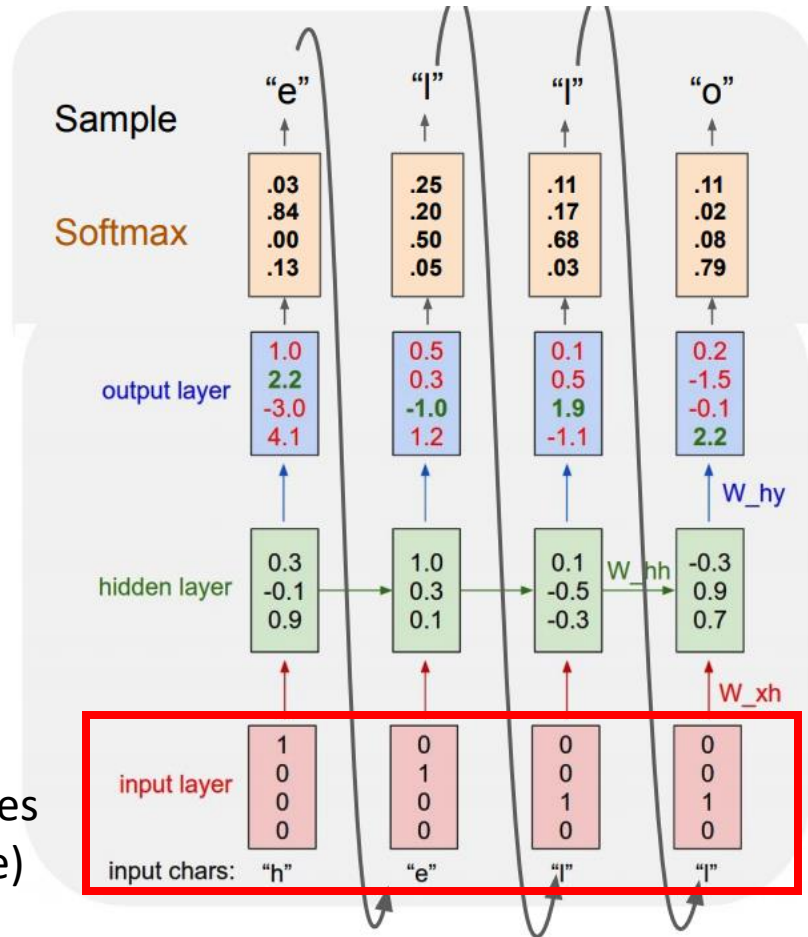


character features

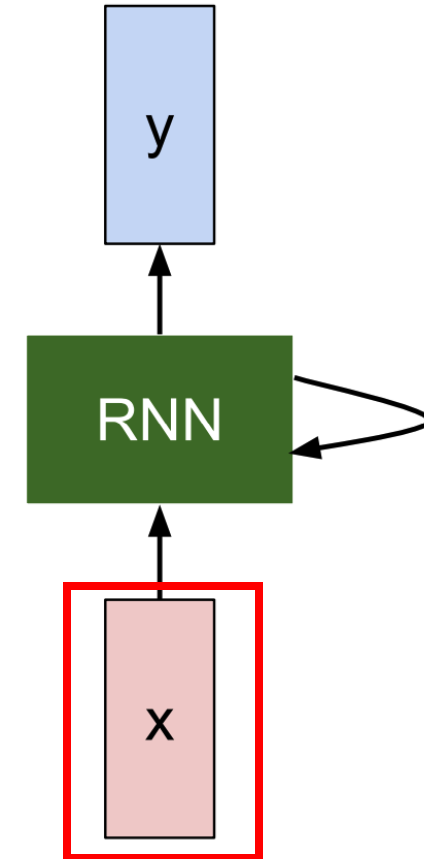


Character-level language model

- Vocabulary: {h, e, l, o}

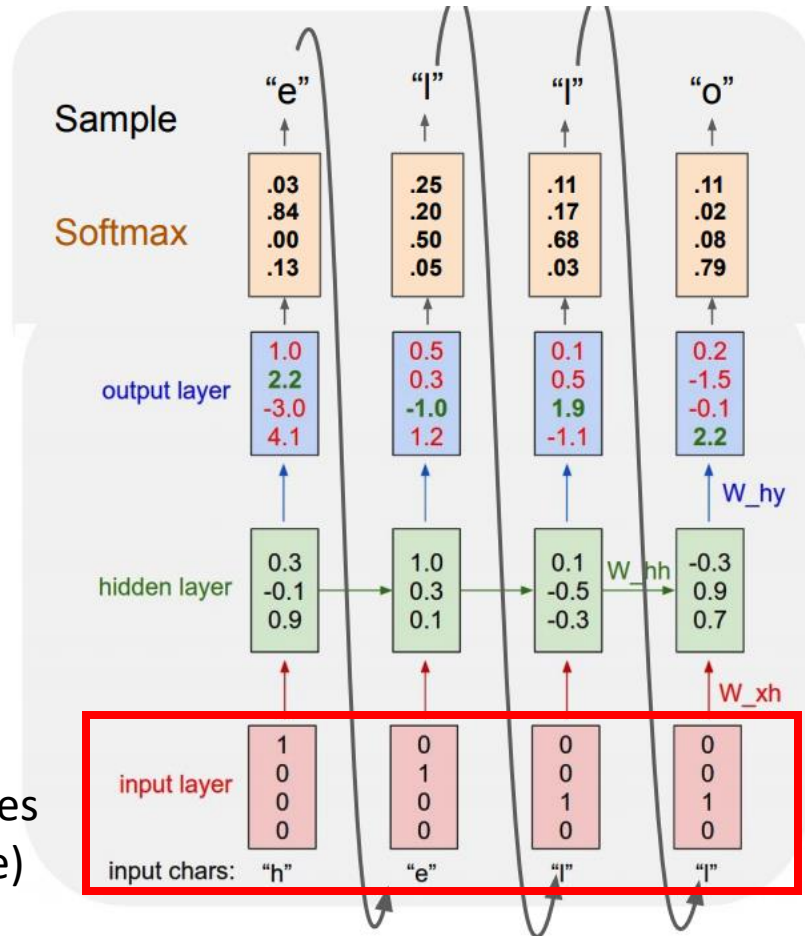


character features
(one-hot encode)

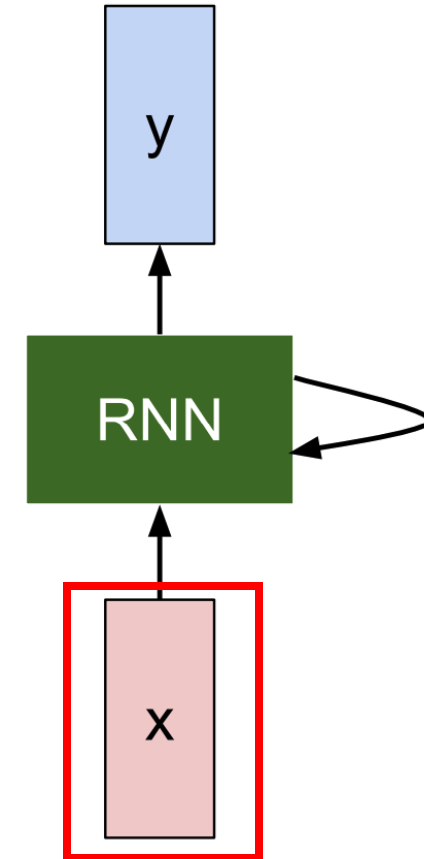
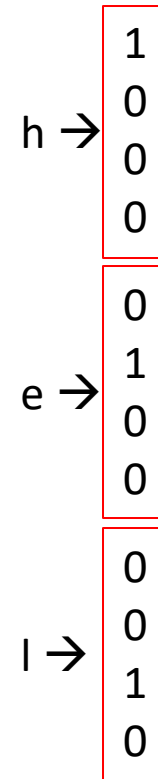


Character-level language model

- Vocabulary: {h, e, l, o}

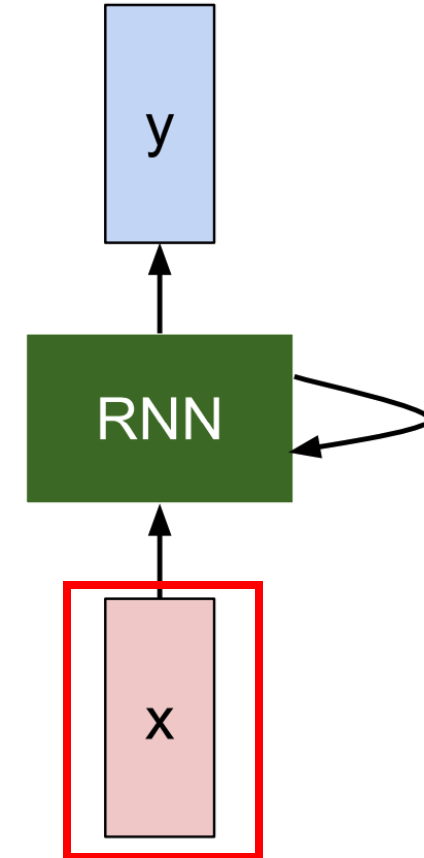
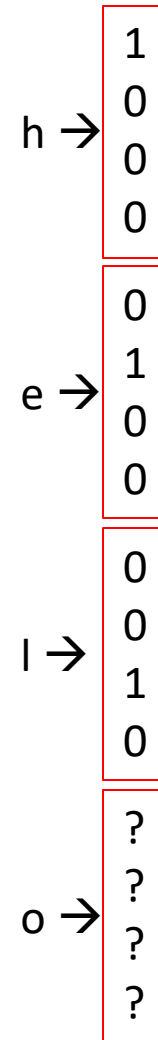
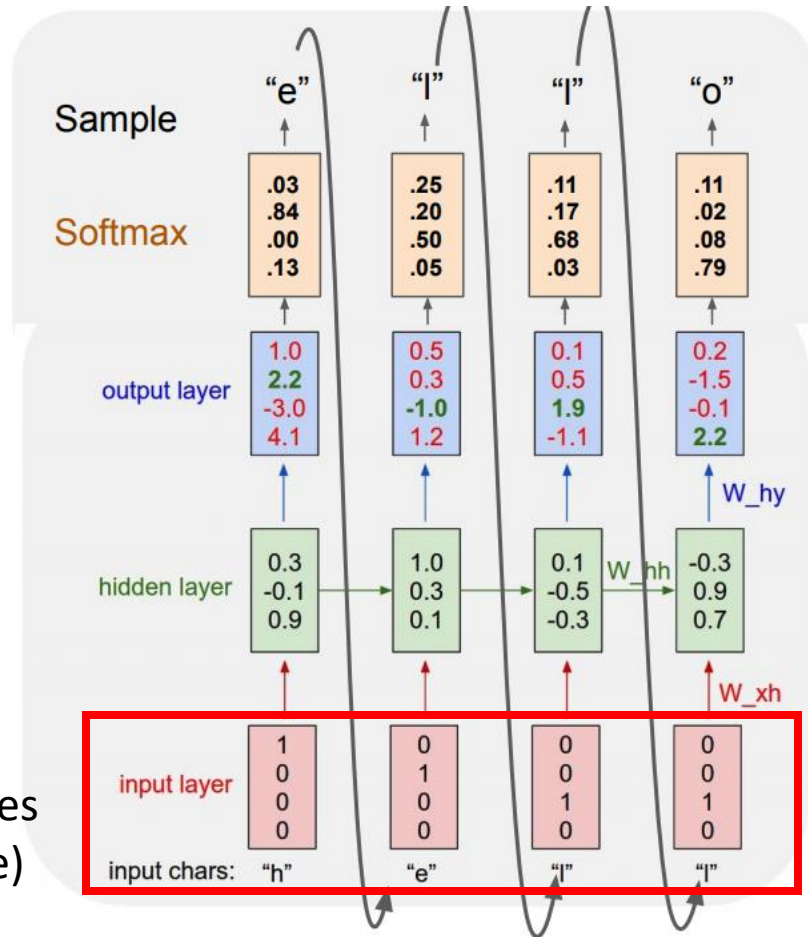


character features
(one-hot encode)



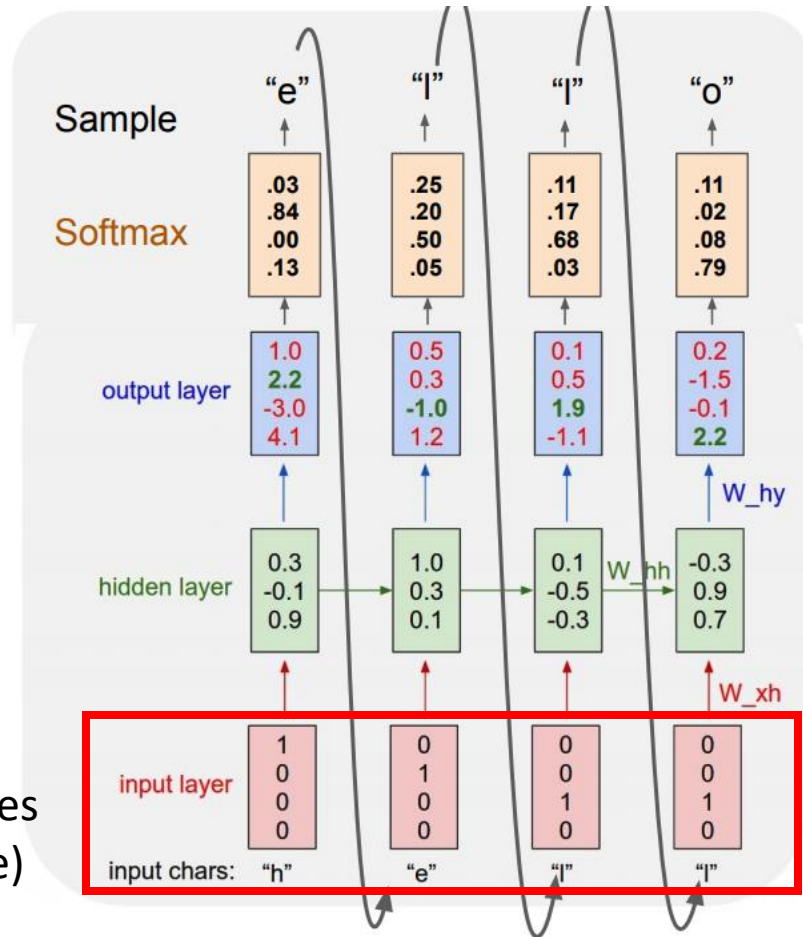
Character-level language model

- Vocabulary: {h, e, l, o}

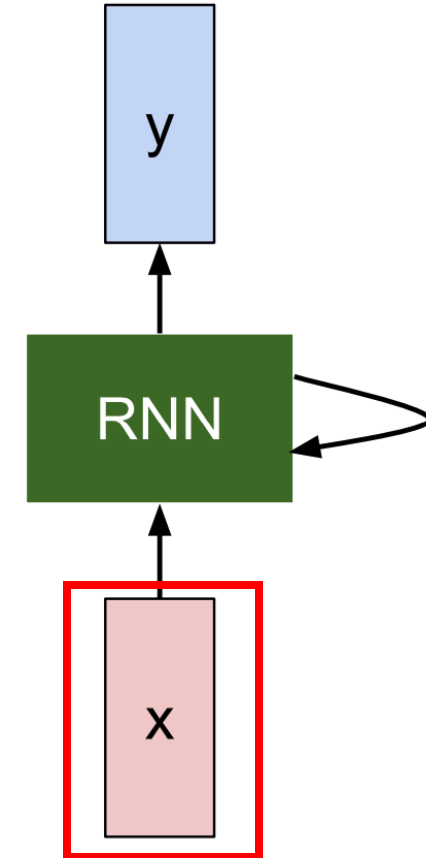
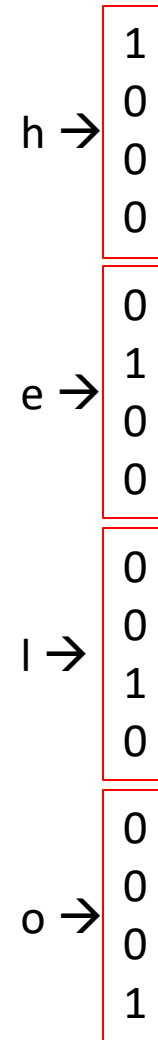


Character-level language model

- Vocabulary: {h, e, l, o}

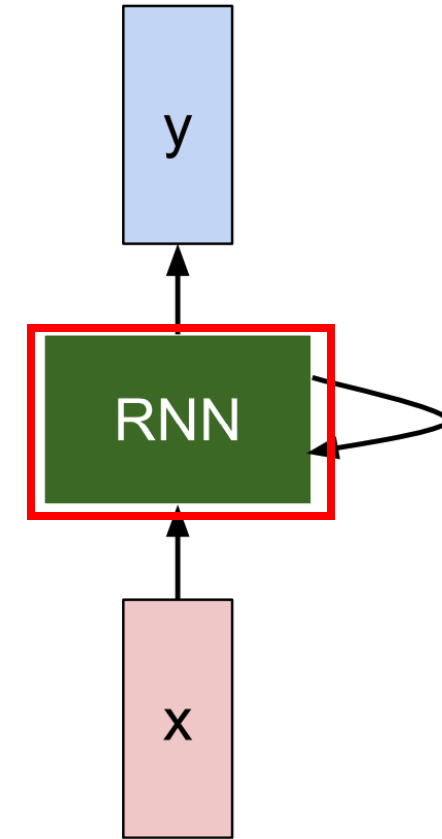
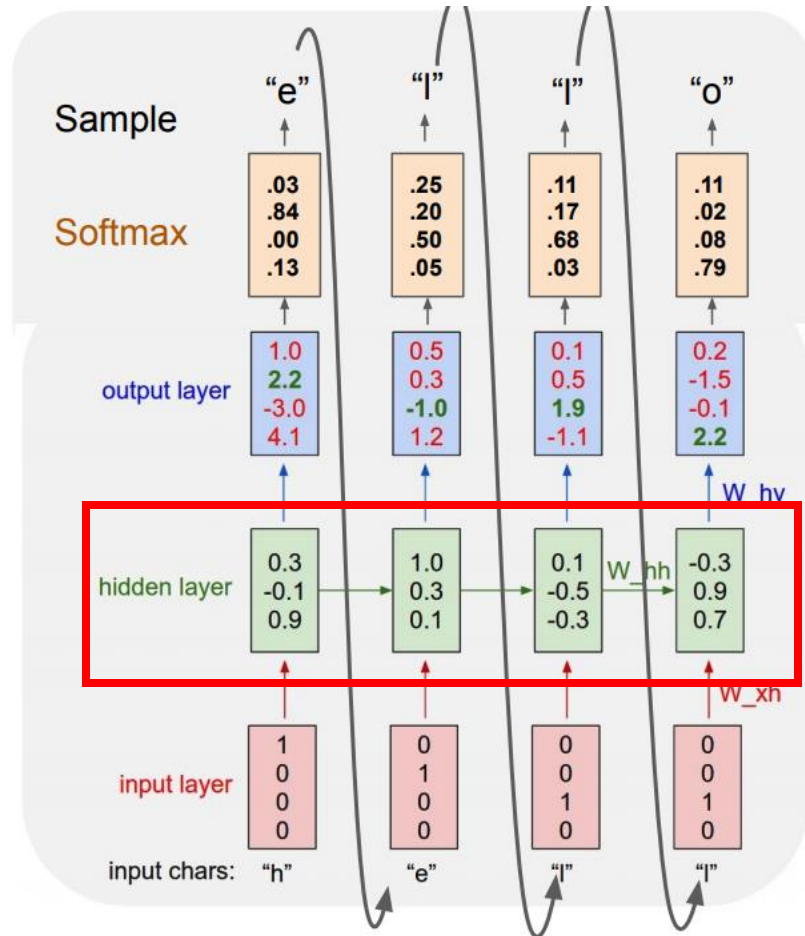


character features
(one-hot encode)



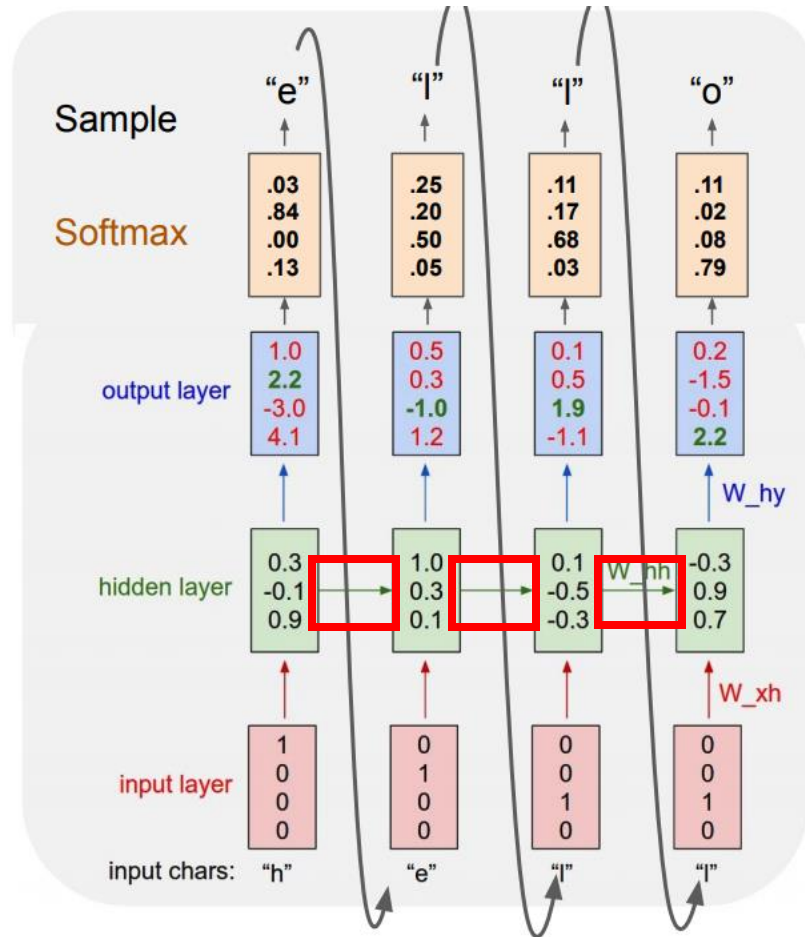
Character-level language model

- Vocabulary: {h, e, l, o}

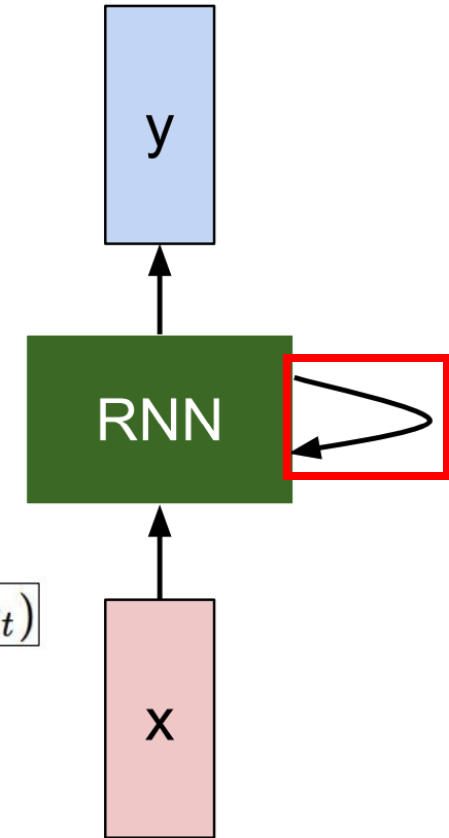


Character-level language model

- Vocabulary: {h, e, l, o}

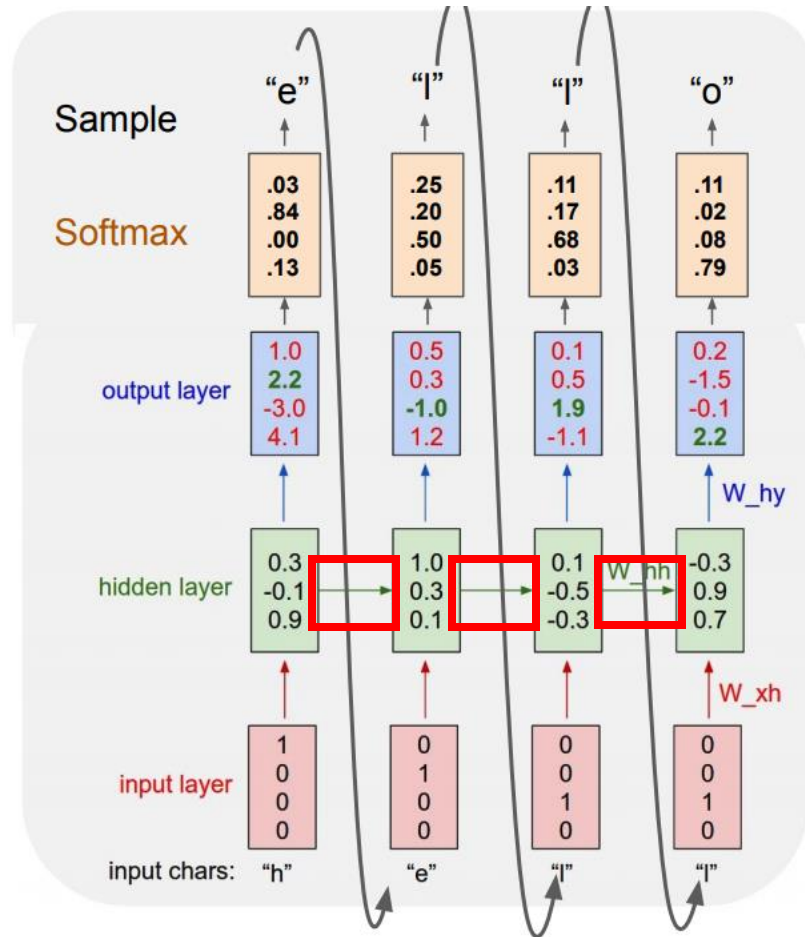


$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t)$$

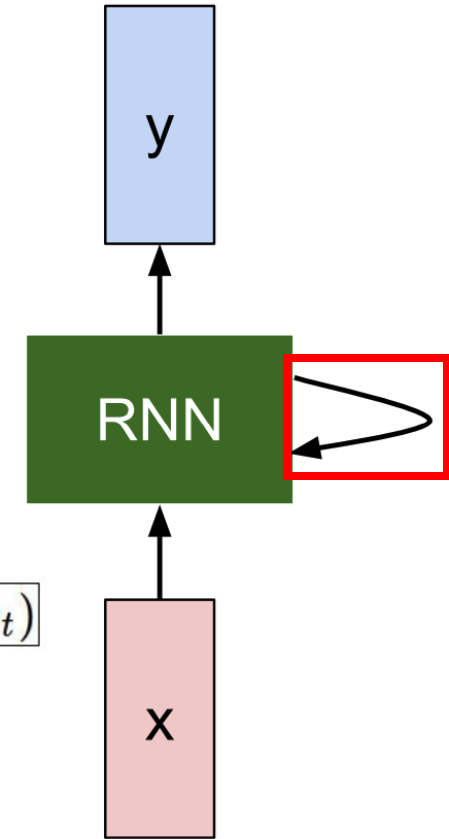


Character-level language model

- Vocabulary: {h, e, l, o}

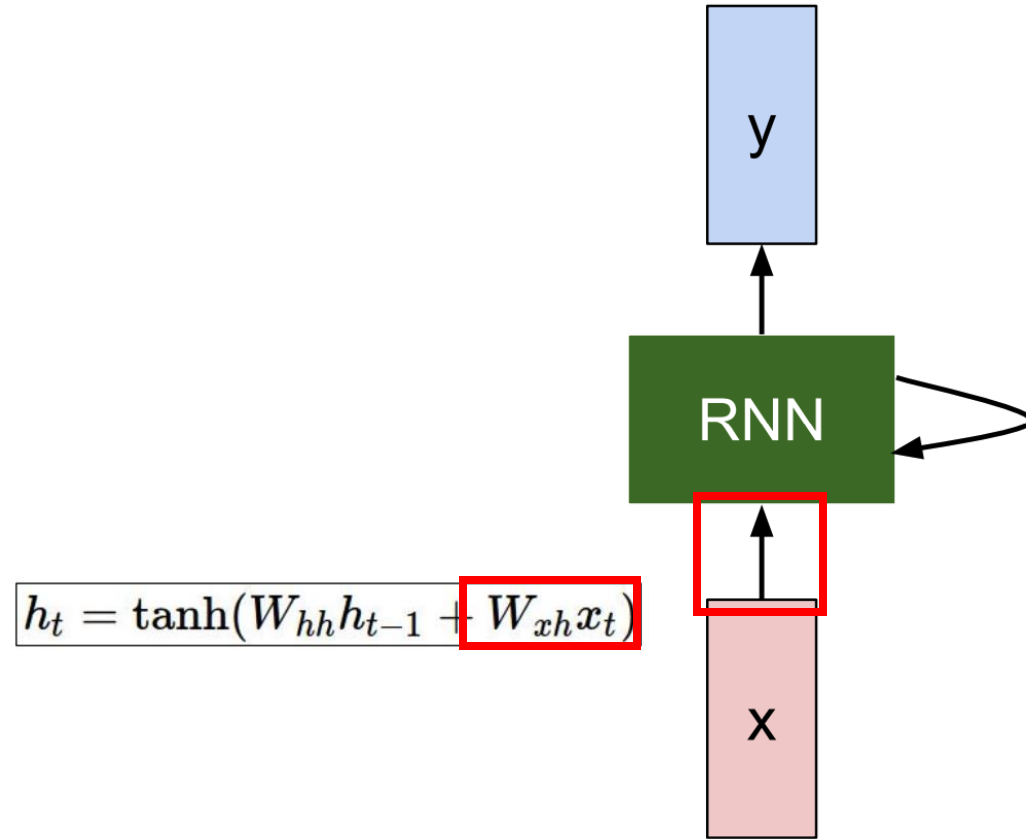
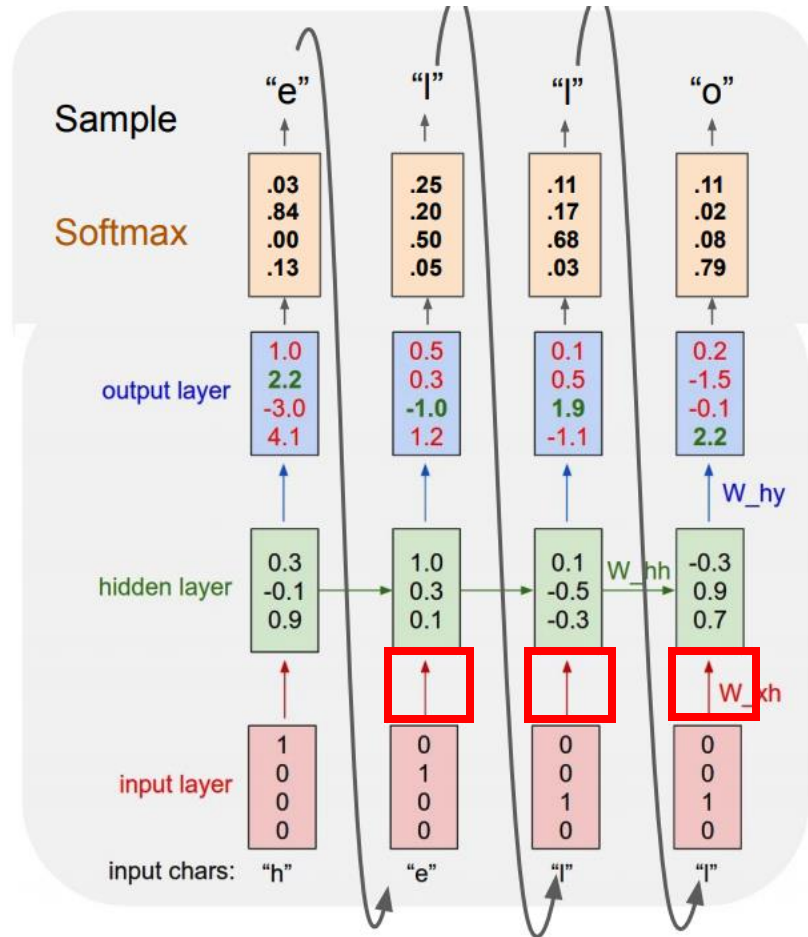


$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t)$$



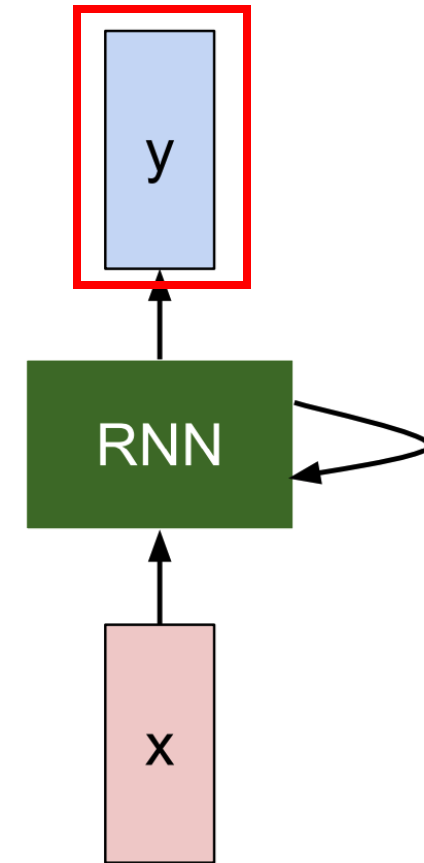
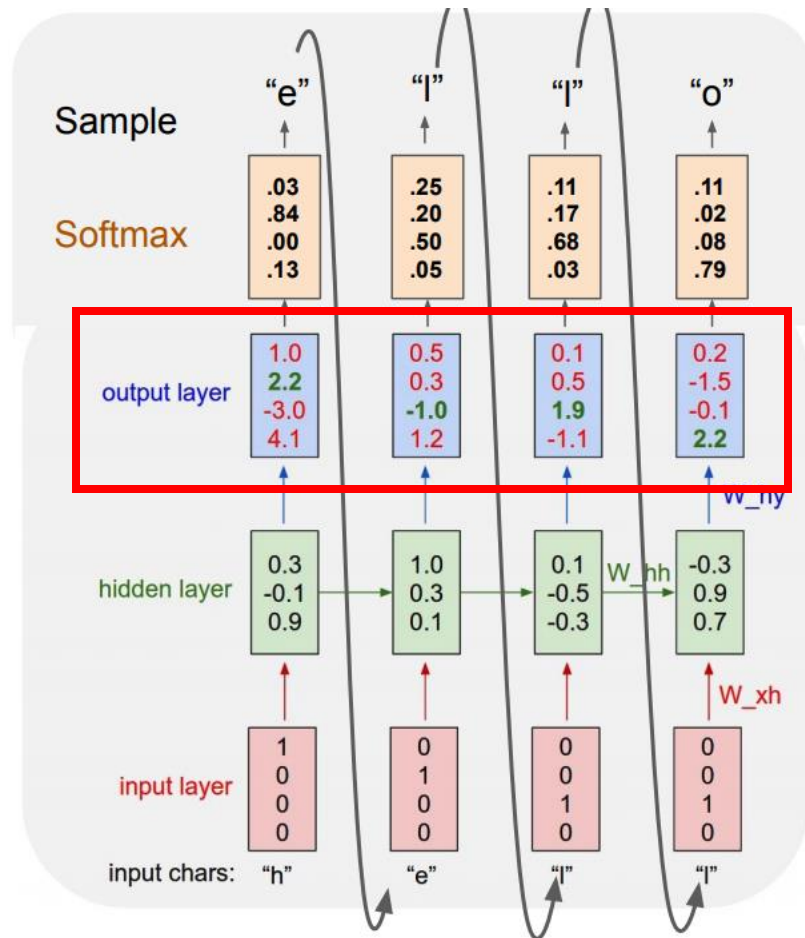
Character-level language model

- Vocabulary: {h, e, l, o}



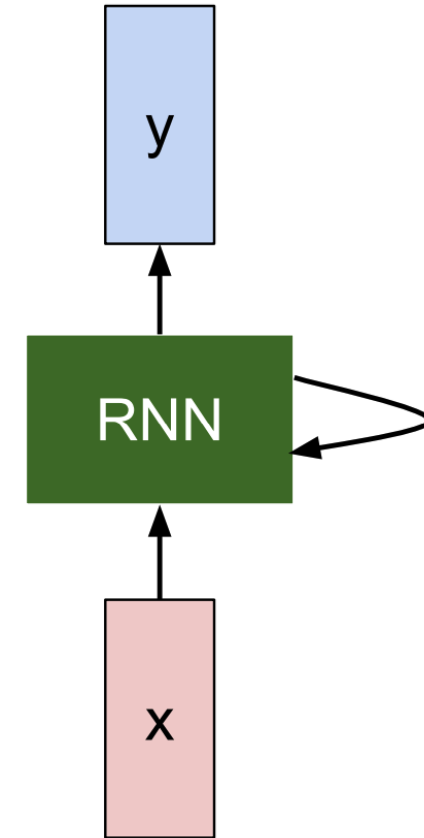
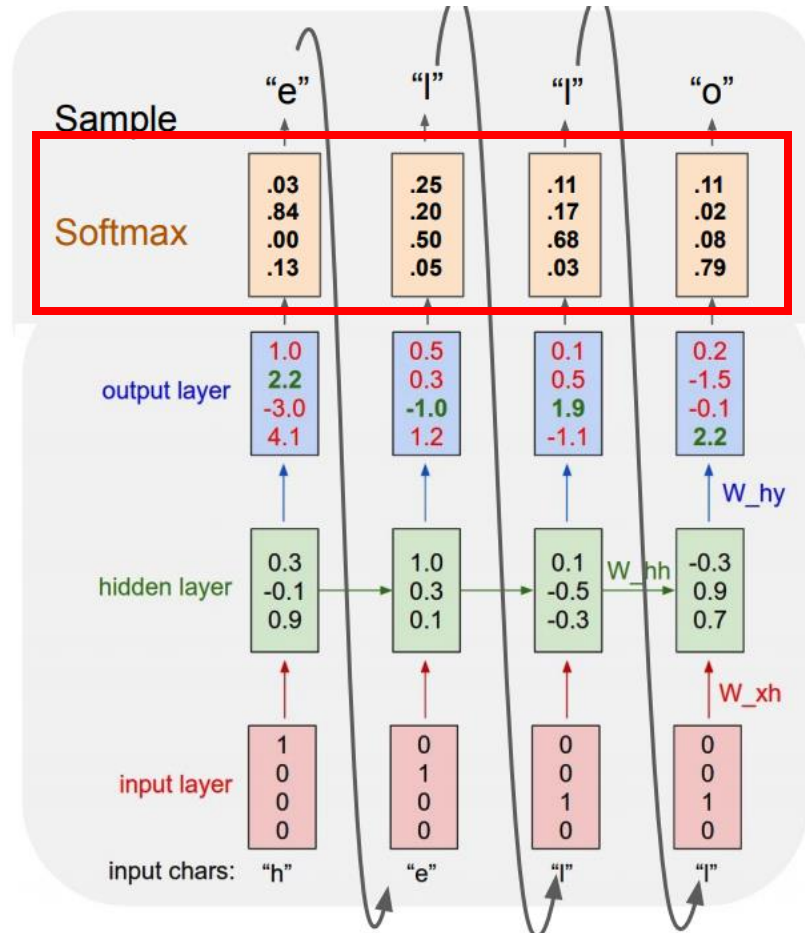
Character-level language model

- Vocabulary: {h, e, l, o}



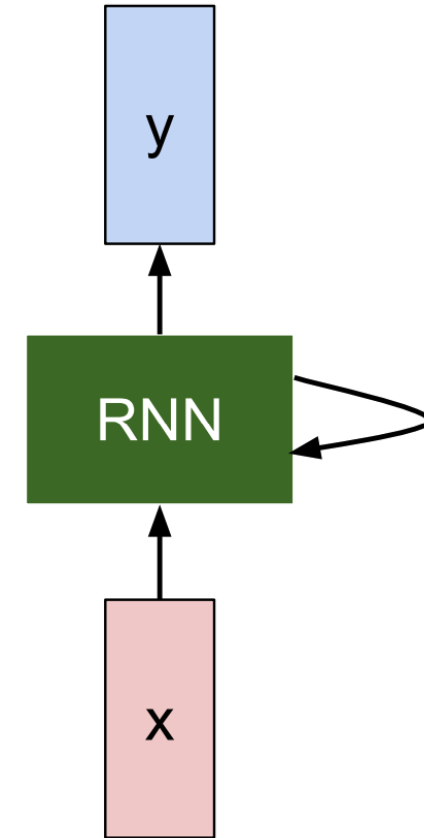
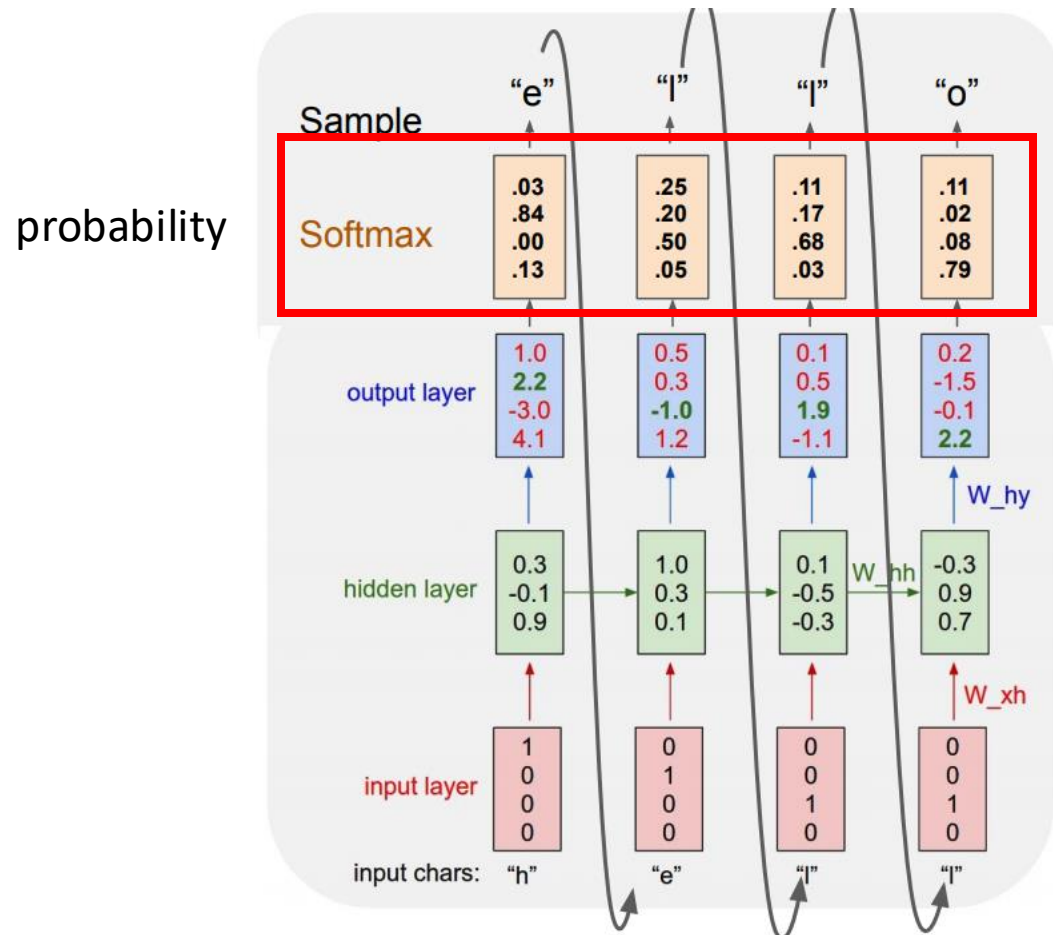
Character-level language model

- Vocabulary: {h, e, l, o}



Character-level language model

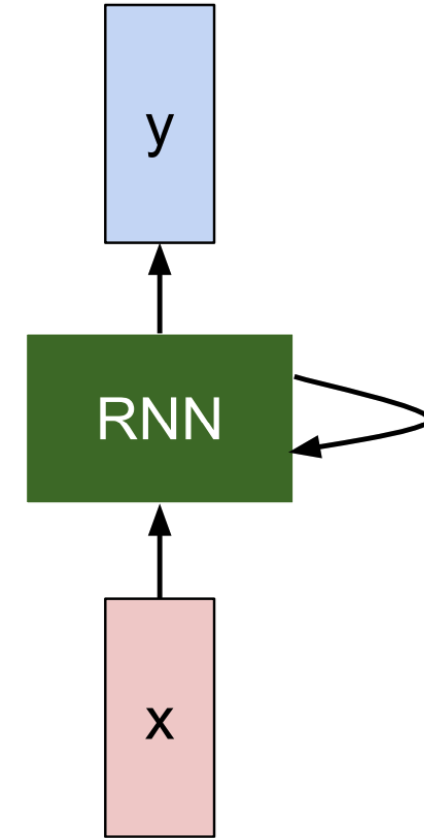
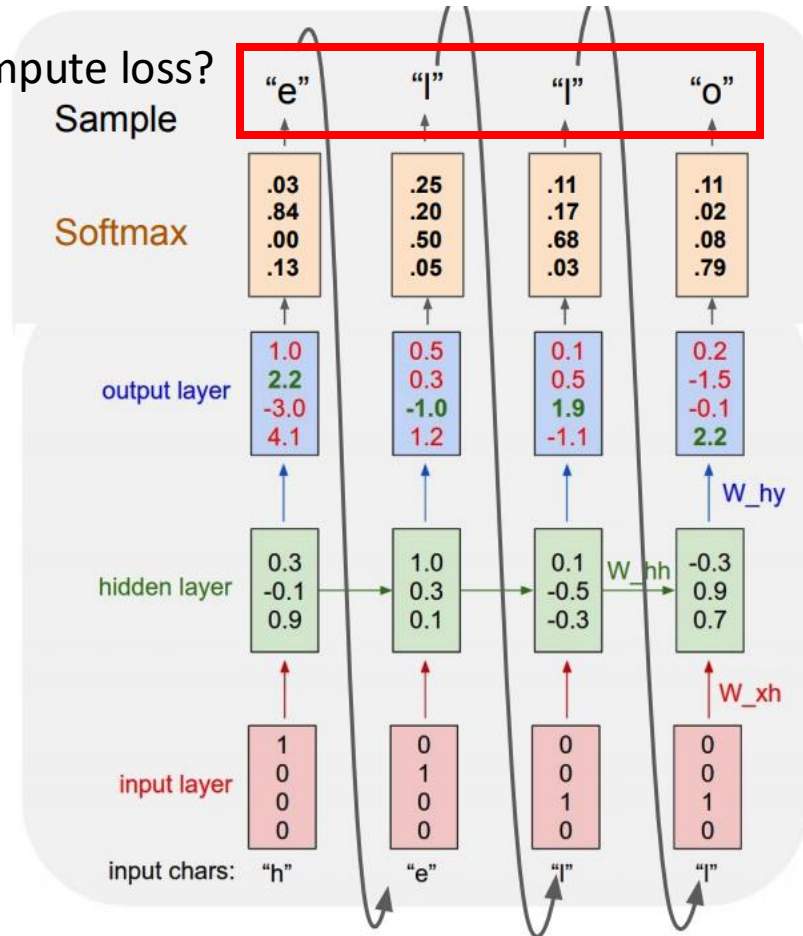
- Vocabulary: {h, e, l, o}



Character-level language model

- Vocabulary: {h, e, l, o}

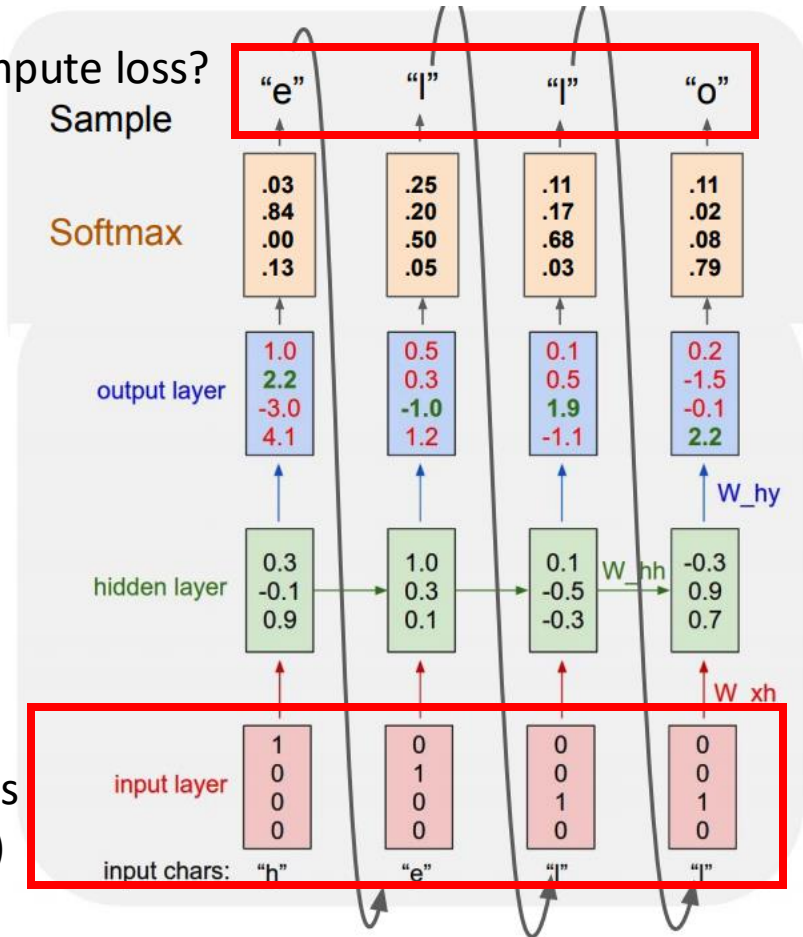
Q: How to compute loss?



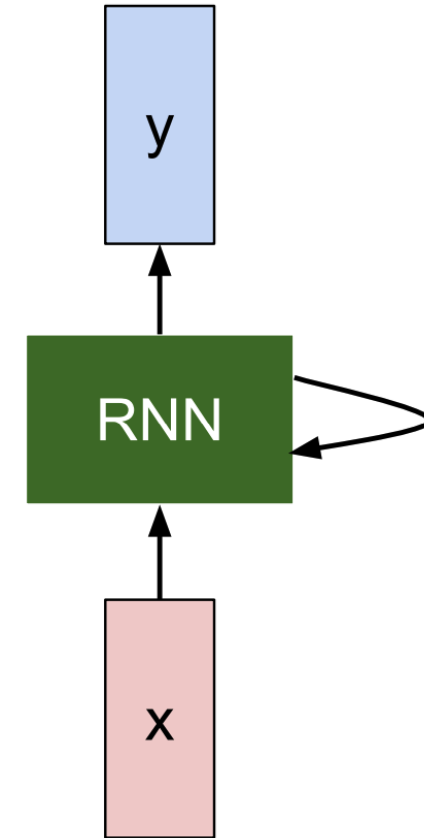
Character-level language model

- Vocabulary: {h, e, l, o}

Q: How to compute loss?



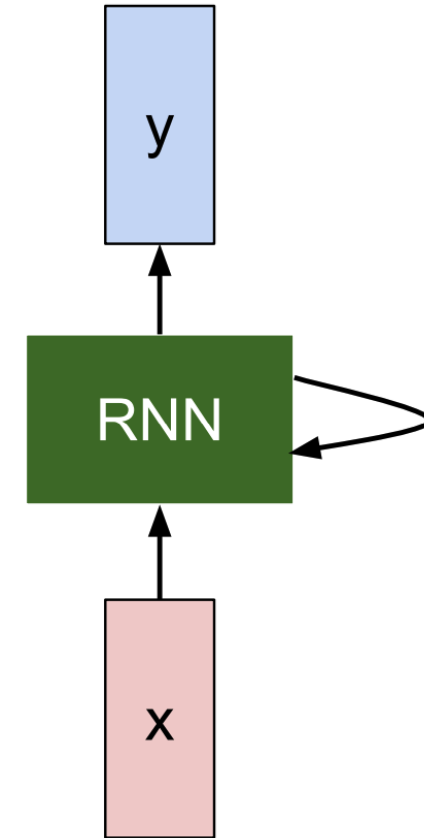
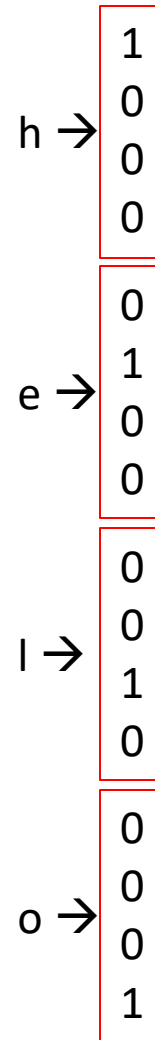
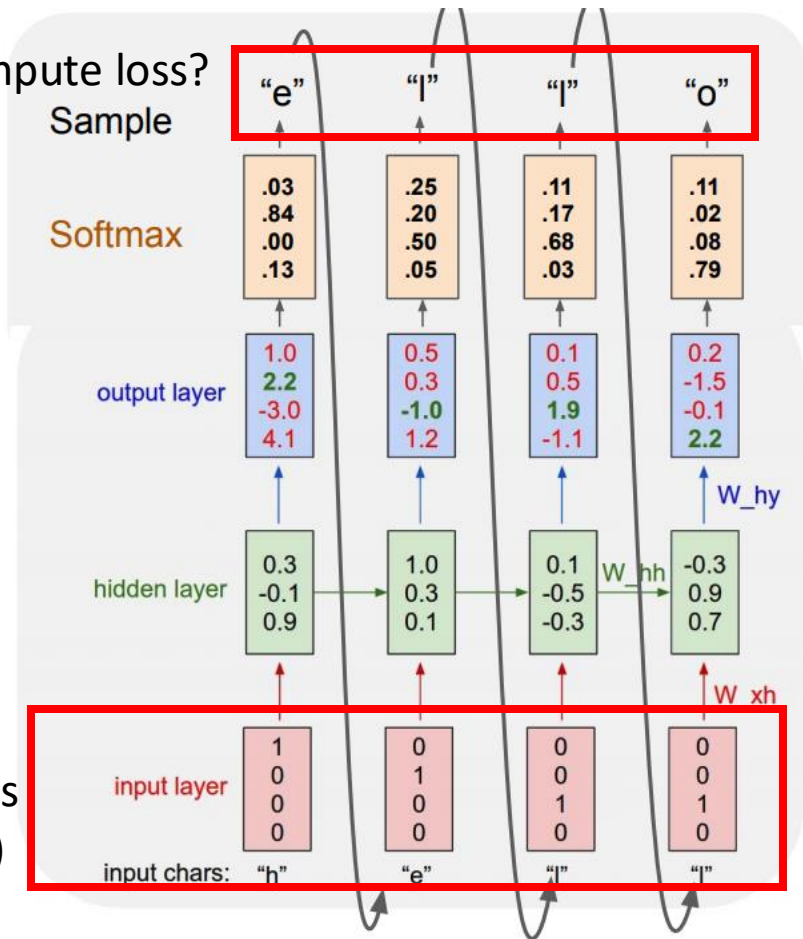
character features
(one-hot encode)



Character-level language model

- Vocabulary: {h, e, l, o}

Q: How to compute loss?

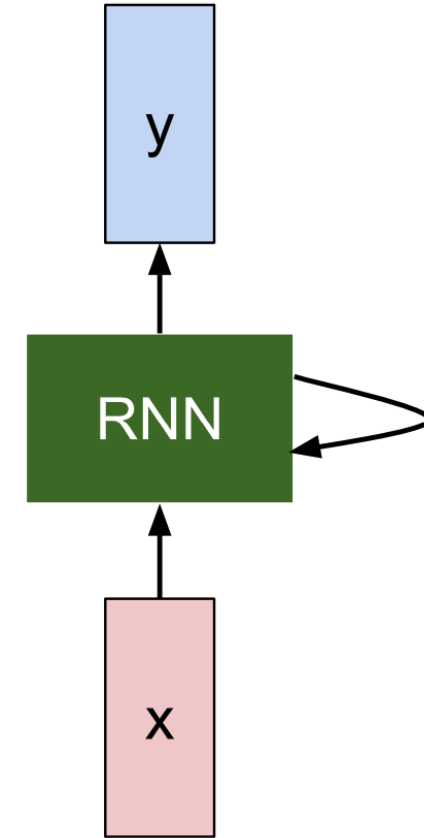
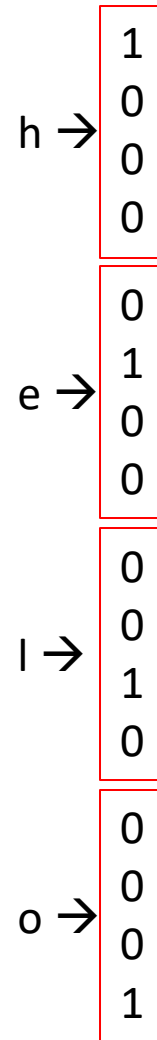
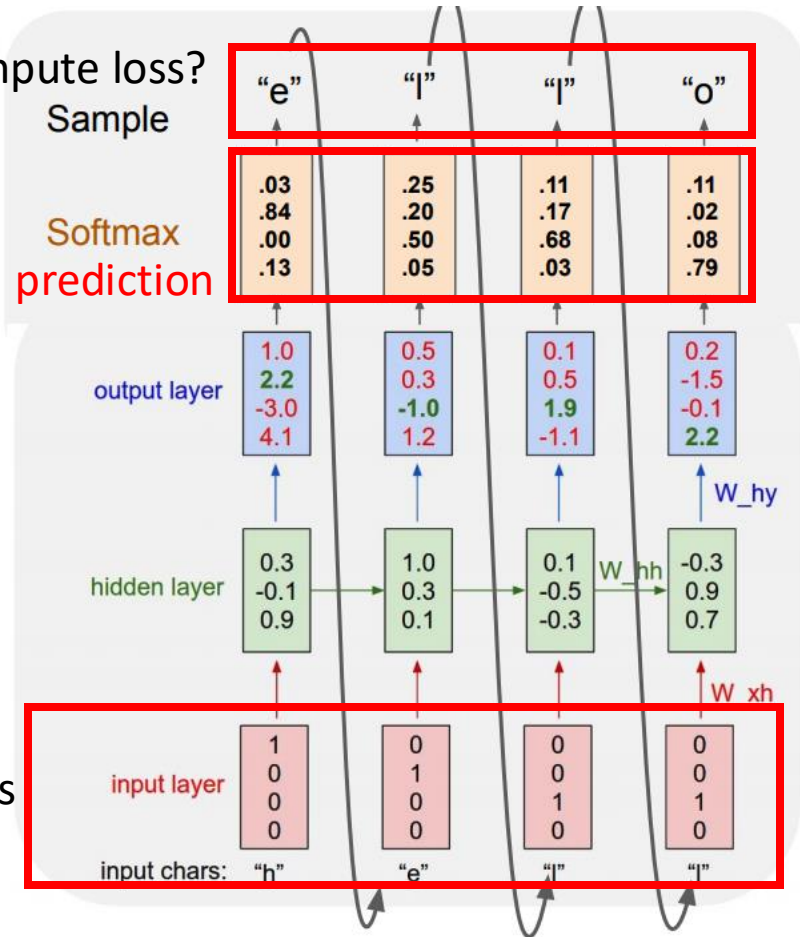


character features
(one-hot encode)

Character-level language model

- Vocabulary: {h, e, l, o}

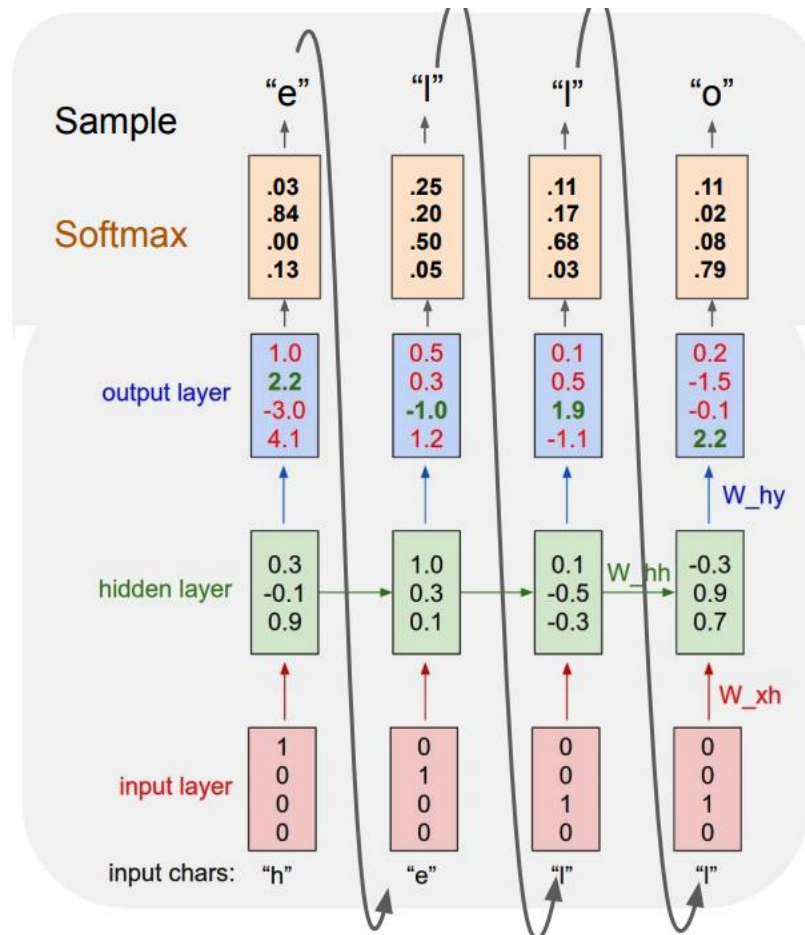
Q: How to compute loss?



character features
(one-hot encode)

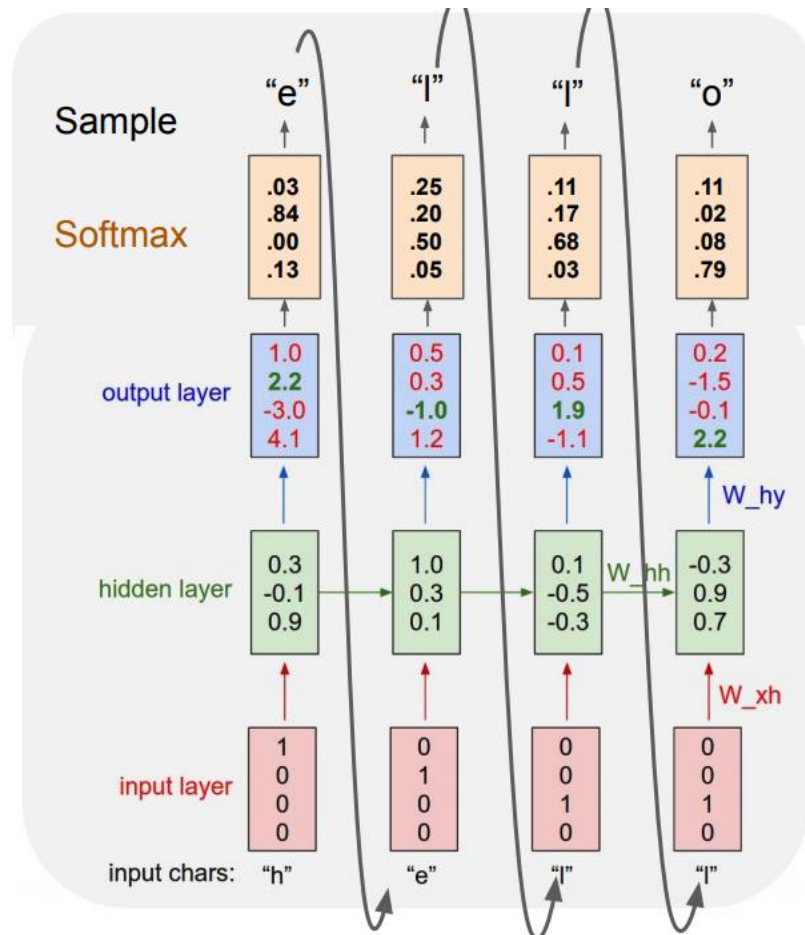
Character-level language model

- Vocabulary: {h, e, l, o}



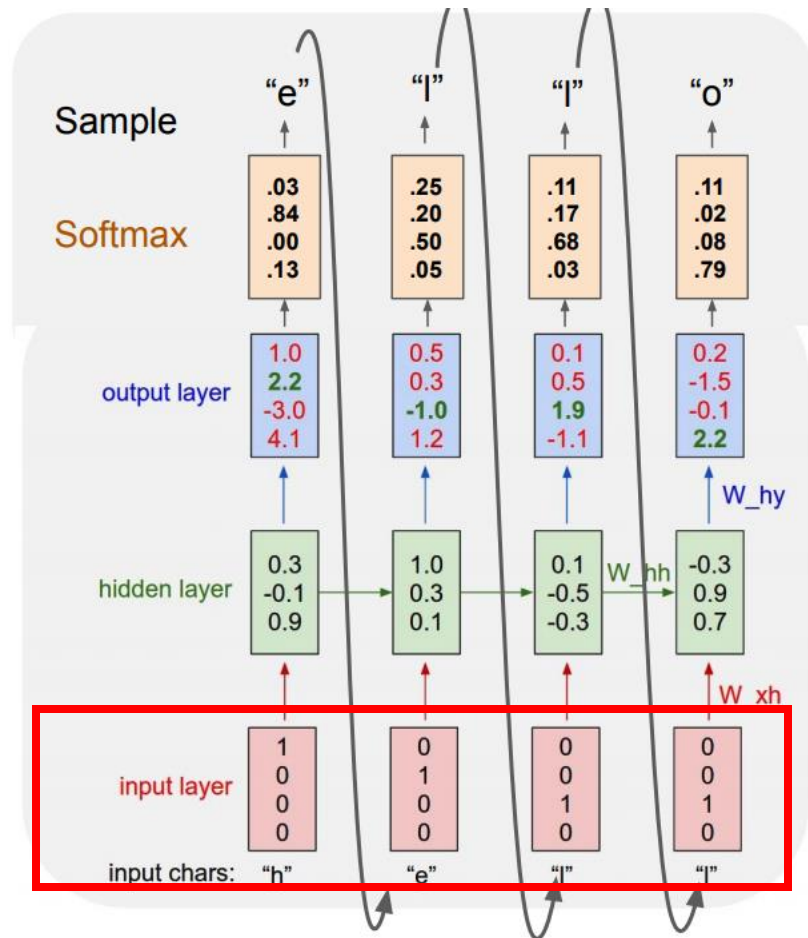
Word-level language model

- Vocabulary: {h, e, l, o} \rightarrow {ant, and, ..., network, ..., zoo}



Word-level language model

- Vocabulary: {h, e, l, o} \rightarrow {ant, and, ..., network, ..., zoo}



Word-level language model

- Vocabulary: {h, e, l, o} $\xrightarrow{\text{Change to}}$ {ant, and, ..., network, ..., zoo}

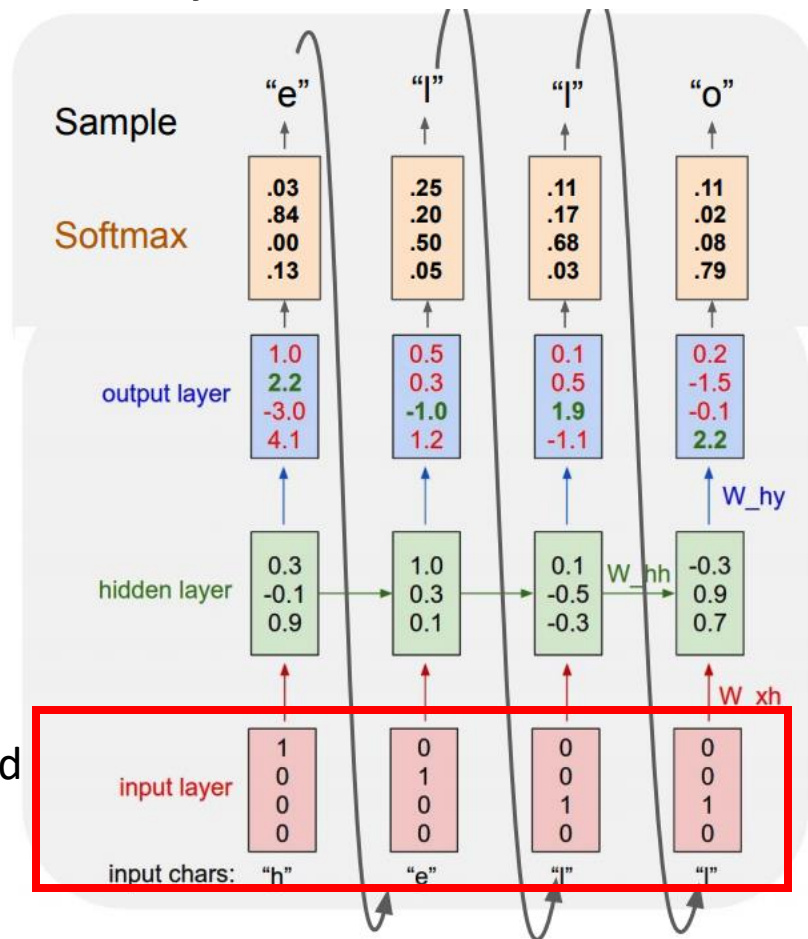


Image captioning

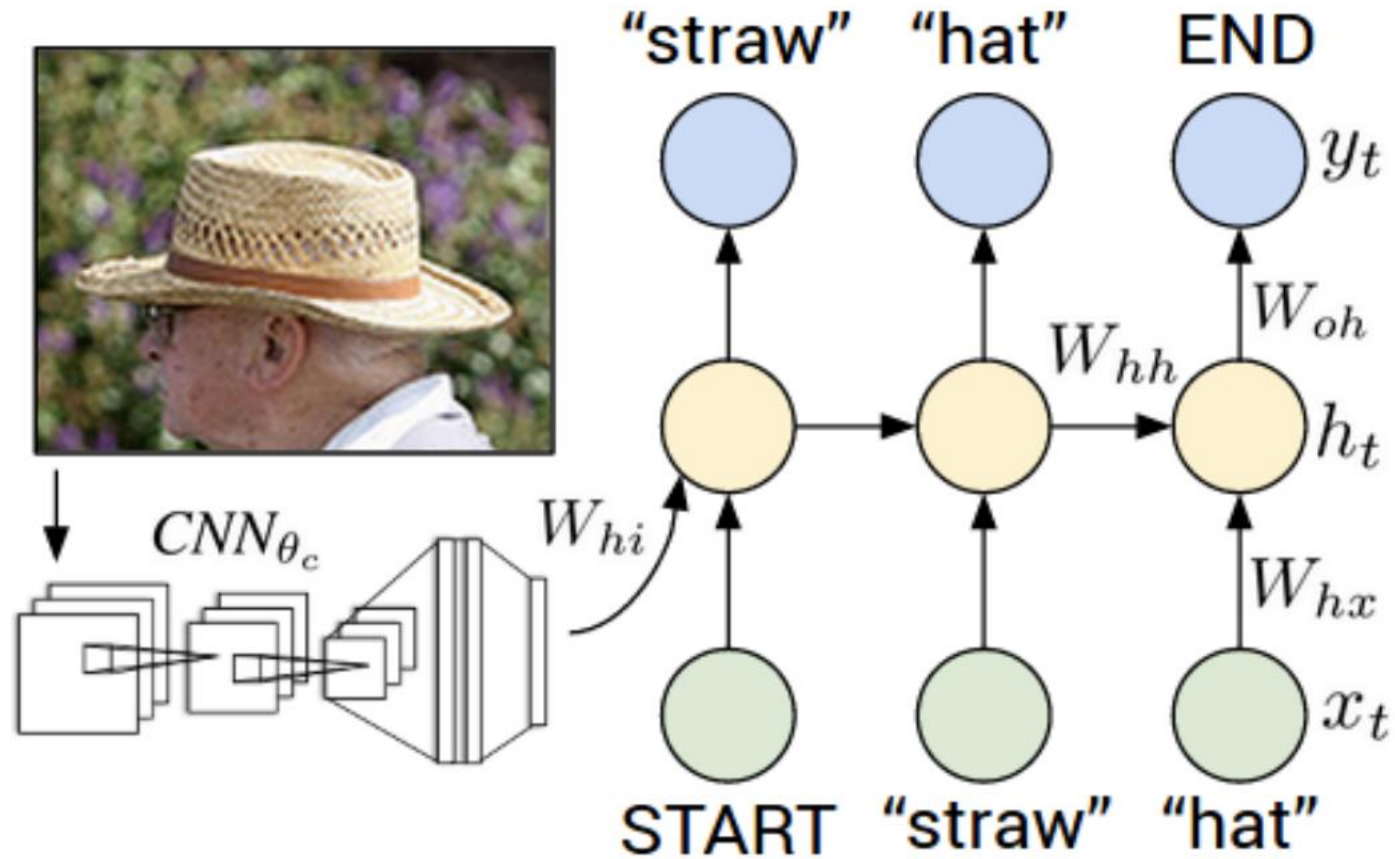


Figure from Karpathy, Andrej, and Li Fei-Fei. "Deep visual-semantic alignments for generating image descriptions." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3128-3137. 2015.

Image captioning

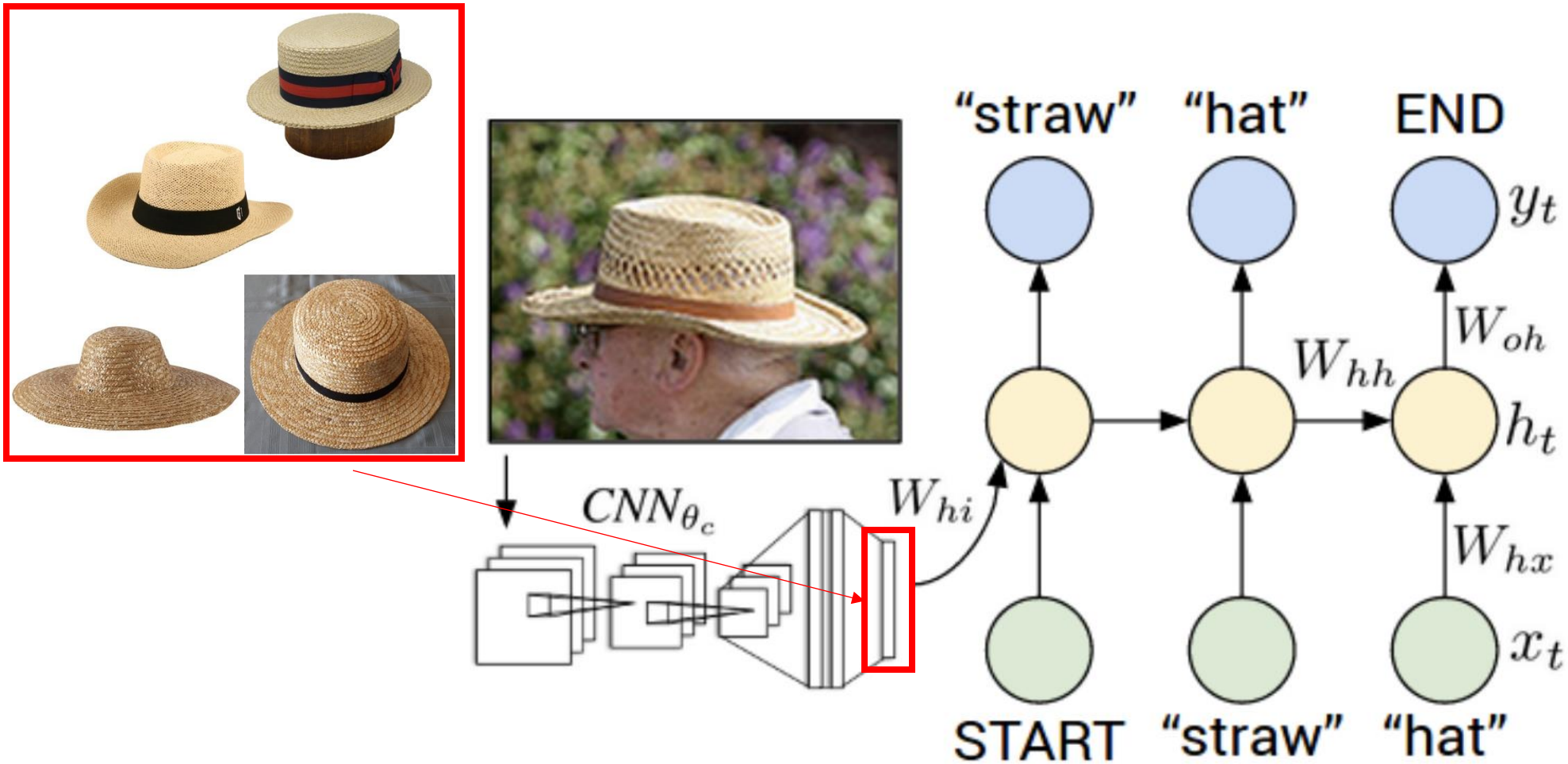


Figure from Karpathy, Andrej, and Li Fei-Fei. "Deep visual-semantic alignments for generating image descriptions." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3128-3137. 2015.

Image captioning

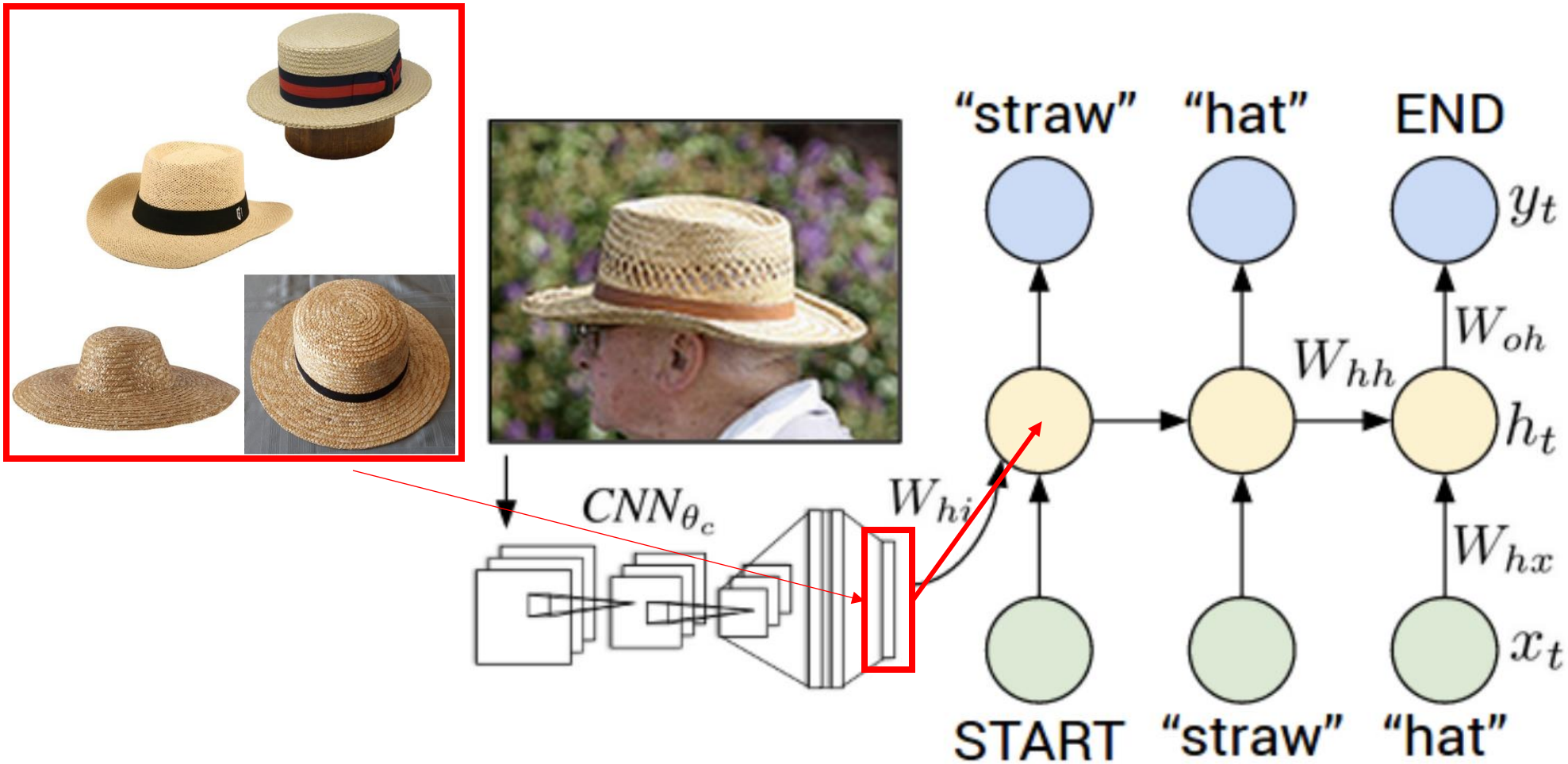
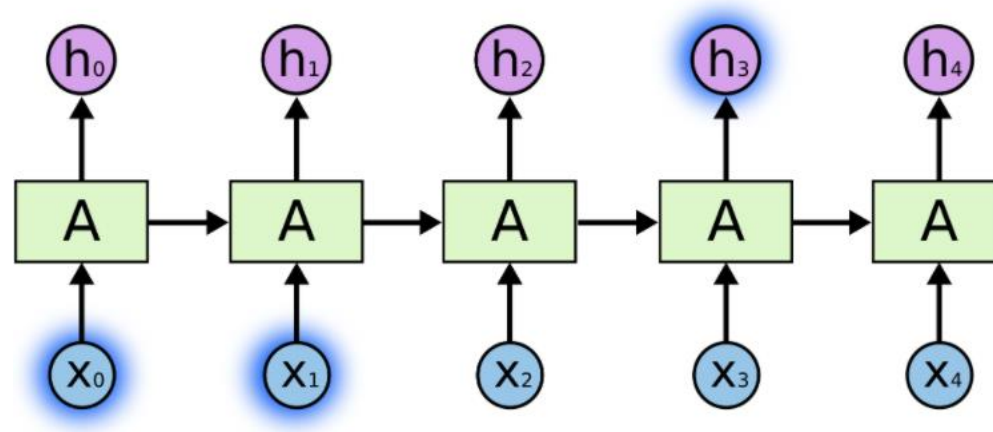


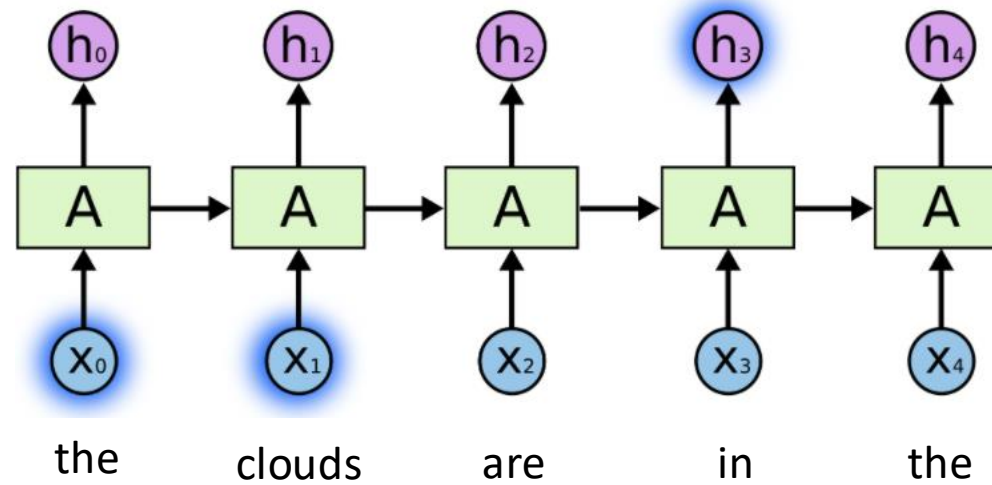
Figure from Karpathy, Andrej, and Li Fei-Fei. "Deep visual-semantic alignments for generating image descriptions." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3128-3137. 2015.

Short-term dependence



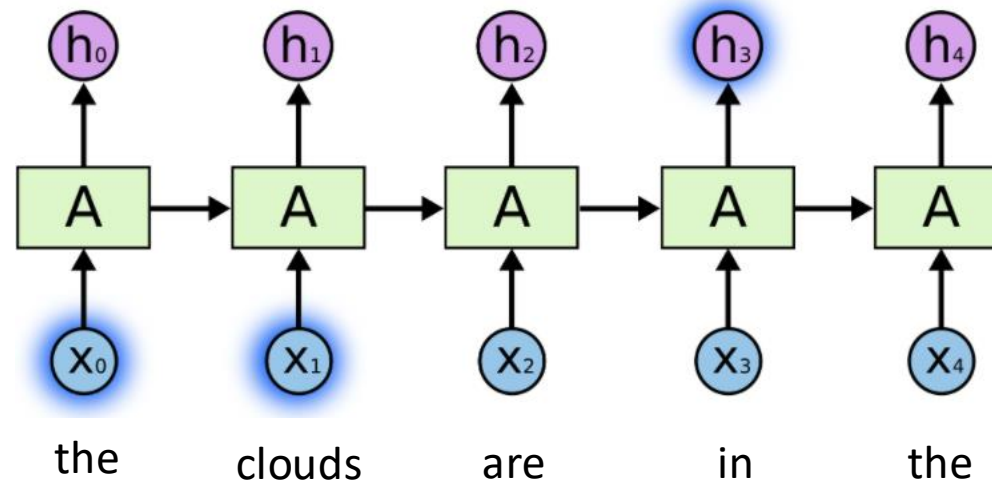
the clouds are in the ???

Short-term dependence



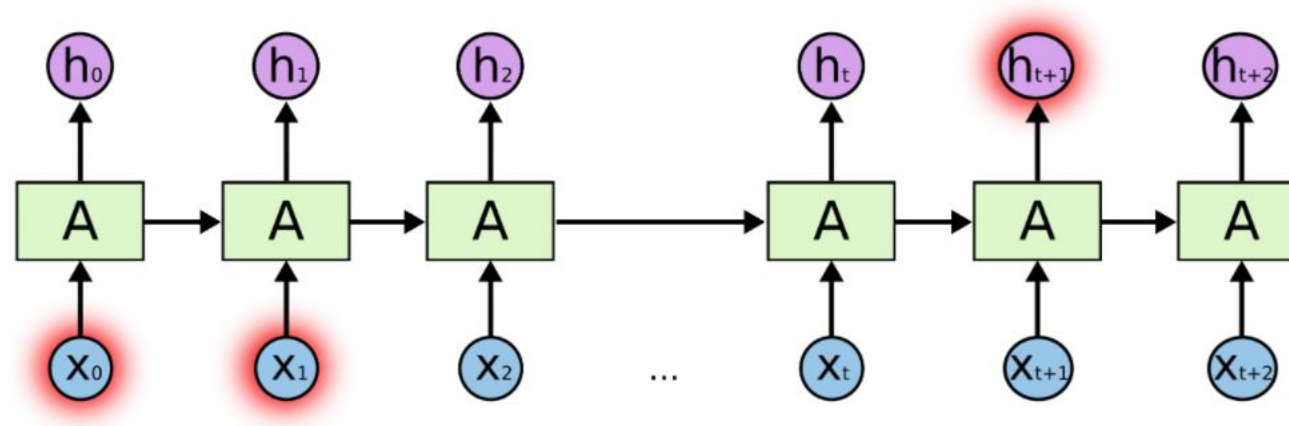
the clouds are in the ???

Short-term dependence



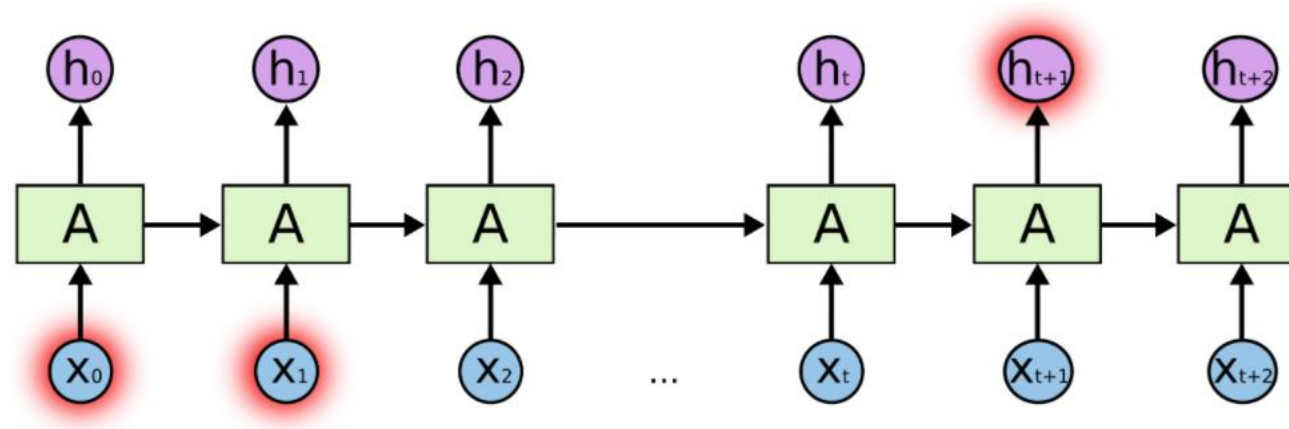
the clouds are in the sky

Long-term dependence



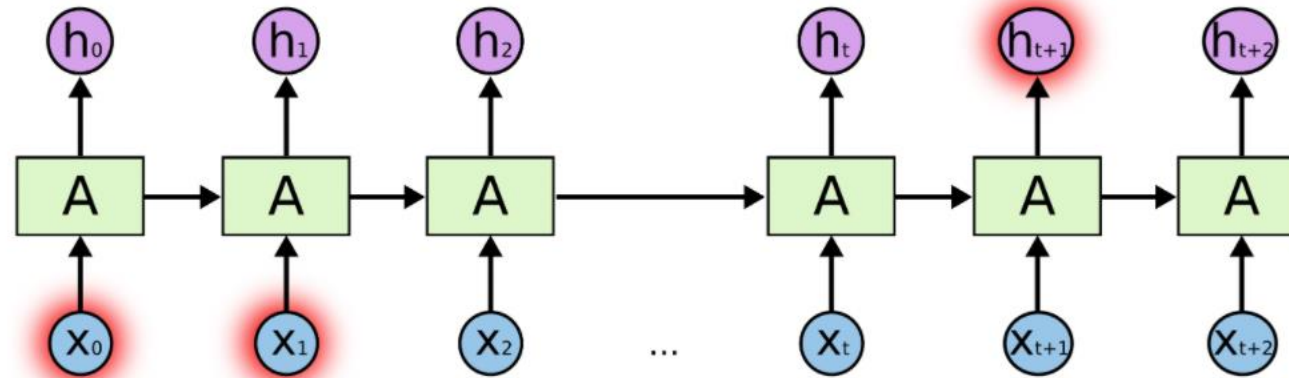
I like this town very much. I started my undergraduate study in 2020 and my major is computer science. I like programming and reading. I usually get up at 7AM and do some exercise. I also go fishing at weekend. I grew up in France.

Long-term dependence



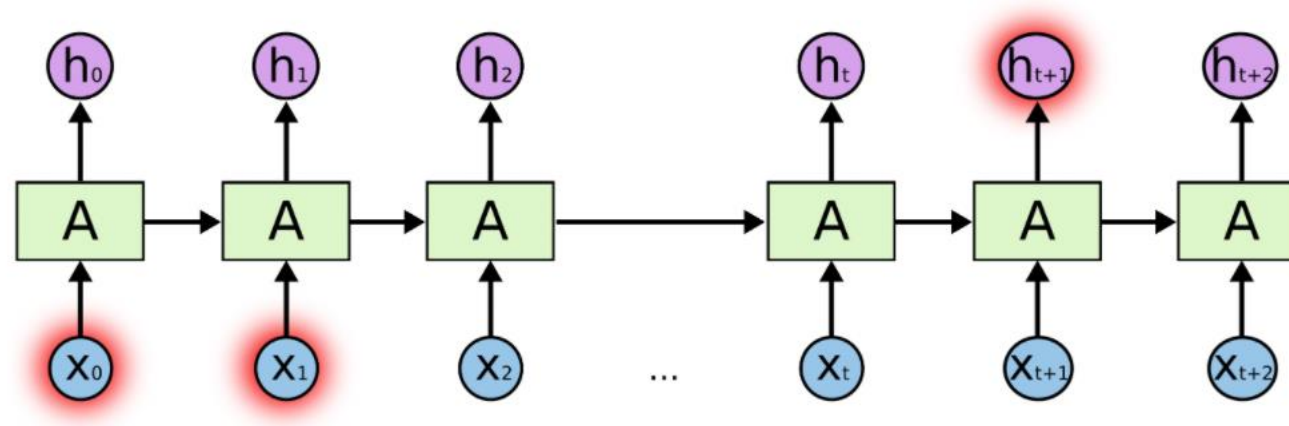
I spent my childhood outdoors.

Long-term dependence



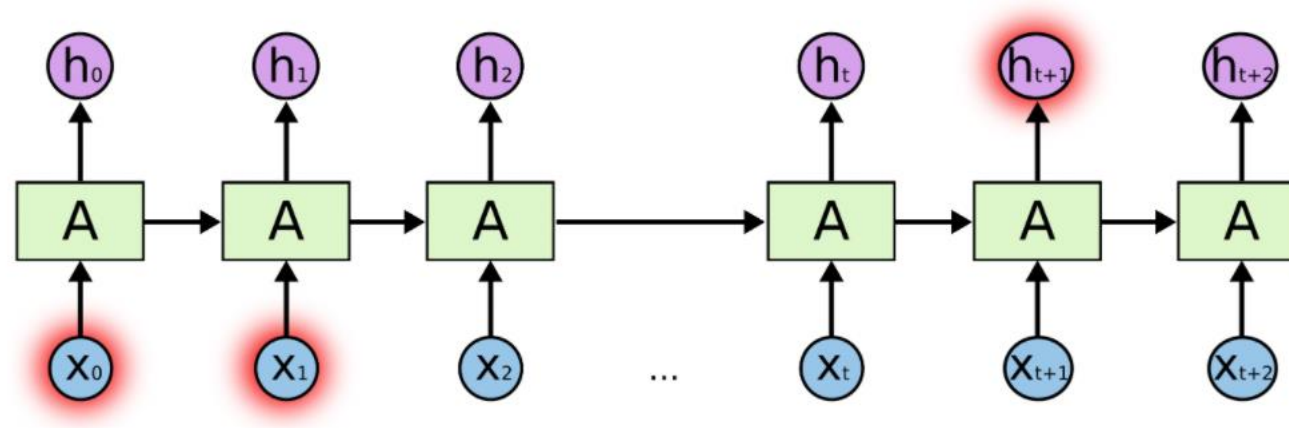
Whether it was riding my bicycle around my neighborhood pretending it was a motorcycle, making mud cakes, going on treasure hunts, making and selling perfume out of strong smelling flowers,

Long-term dependence



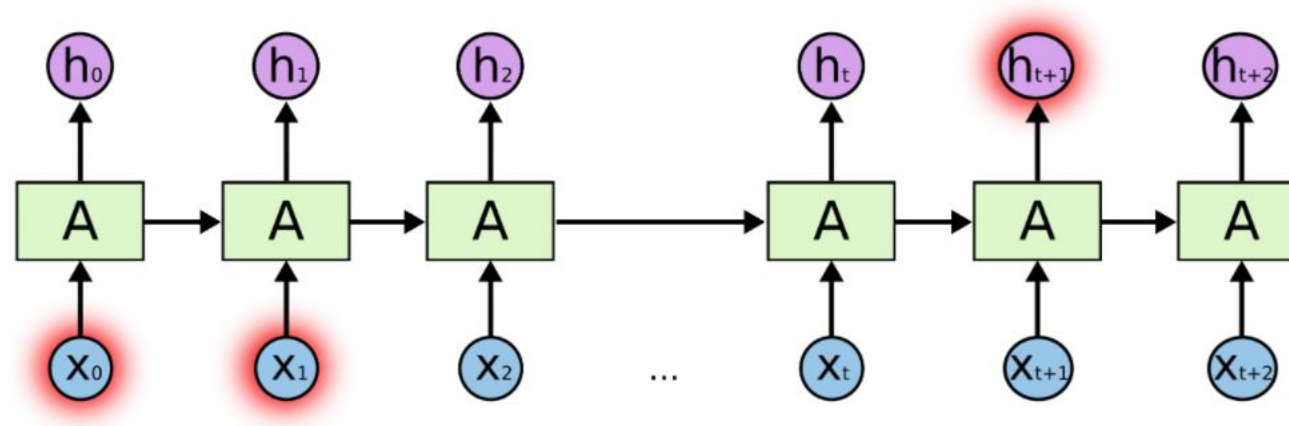
or simply laying on the grass underneath the sun with a soccer ball waiting for someone to come out and play with me, the outdoors was where I spent my childhood and I cannot be more appreciative of it.

Long-term dependence



I speak fluent ???.

Long-term dependence

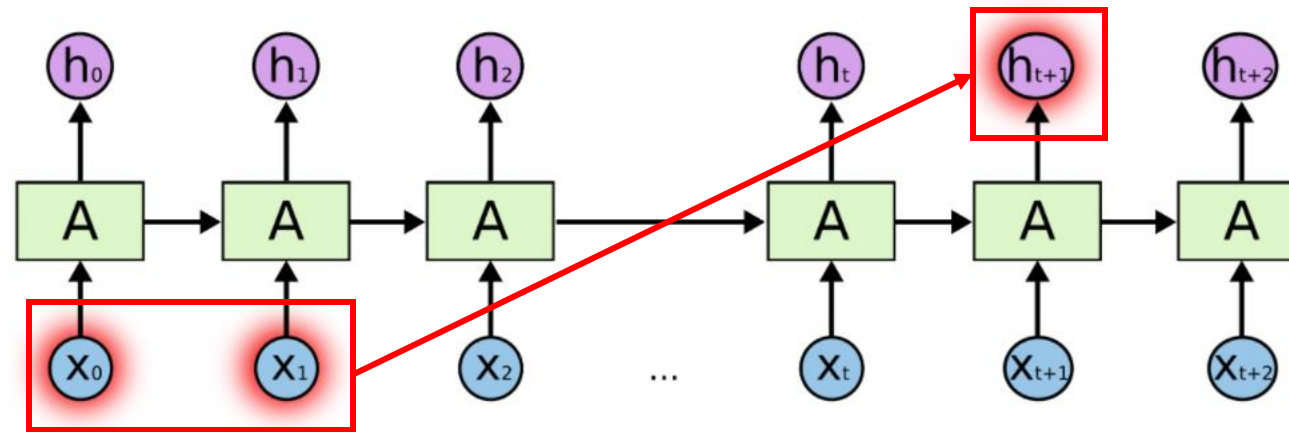


I like this town very much. I started my undergraduate study in 2020 and my major is computer science. I like programming and reading. I usually get up at 7AM and do some exercise. I also go fishing at weekend. I grew up in **France**.

I spent my childhood outdoors. Whether it was riding my bicycle around my neighborhood pretending it was a motorcycle, making mud cakes, going on treasure hunts, making and selling perfume out of strong smelling flowers, or simply laying on the grass underneath the sun with a soccer ball waiting for someone to come out and play with me, the outdoors was where I spent my childhood and I cannot be more appreciative of it.

I speak fluent **French**.

Long-term dependence

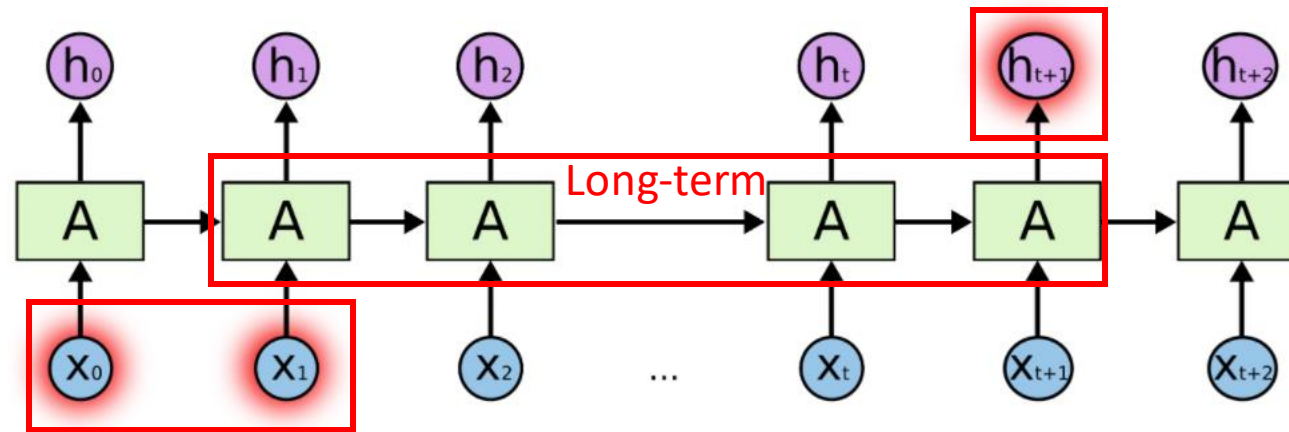


I like this town very much. I started my undergraduate study in 2020 and my major is computer science. I like programming and reading. I usually get up at 7AM and do some exercise. I also go fishing at weekend. I grew up in **France**.

I spent my childhood outdoors. Whether it was riding my bicycle around my neighborhood pretending it was a motorcycle, making mud cakes, going on treasure hunts, making and selling perfume out of strong smelling flowers, or simply laying on the grass underneath the sun with a soccer ball waiting for someone to come out and play with me, the outdoors was where I spent my childhood and I cannot be more appreciative of it.

I speak fluent **French**.

Long-term dependence



I like this town very much. I started my undergraduate study in 2020 and my major is computer science. I like programming and reading. I usually get up at 7AM and do some exercise. I also go fishing at weekend. I grew up in **France**.

I spent my childhood outdoors. Whether it was riding my bicycle around my neighborhood pretending it was a motorcycle, making mud cakes, going on treasure hunts, making and selling perfume out of strong smelling flowers, or simply laying on the grass underneath the sun with a soccer ball waiting for someone to come out and play with me, the outdoors was where I spent my childhood and I cannot be more appreciative of it.

I speak fluent **French**.