

# DEEP LEARNING WORKSHOP

Dublin City University  
21-22 May 2018



#InsightDL2018

## Case study III

# Skipping and Repeating Samples in RNNs



Xavier Giro-i-Nieto

xavier.giro@upc.edu

Associate Professor

Intelligent Data Science and Artificial Intelligence Center  
Universitat Politècnica de Catalunya (UPC)

# Our team @ UPC IDEAI Barcelona



Xavi  
Giró



Amaia  
Salvador



Victor  
Campos



Miriam  
Bellver



Marc  
Assens



Amanda  
Duarte



Dani  
Fojo



Görkem  
Çamli



Santiago  
Pascual



Marta R.  
Costa.jussà



Ferran  
Marqués



Jordi  
Torres  
(BSC)



Antonio  
Bonafonte



Akis  
Linardos



Sandra  
Roca



Miquel  
LLobet

# Our partners @ academia



Kevin  
McGuinness



Noel E.  
O'Connor



Cathal  
Gurrin



Eva  
Mohedano



Shih-Fu  
Chang



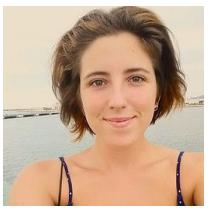
Francisco  
Roldan



Marta  
Coll



Alejandro  
Woodward



Paula  
Gómez



Marc  
Górriz



Carles  
Ventura

# Our team @ industry



Deutsches  
Forschungszentrum  
für Künstliche  
Intelligenz GmbH



Elisenda  
Bou



Dèlia  
Fernández



Sebastian  
Palacio



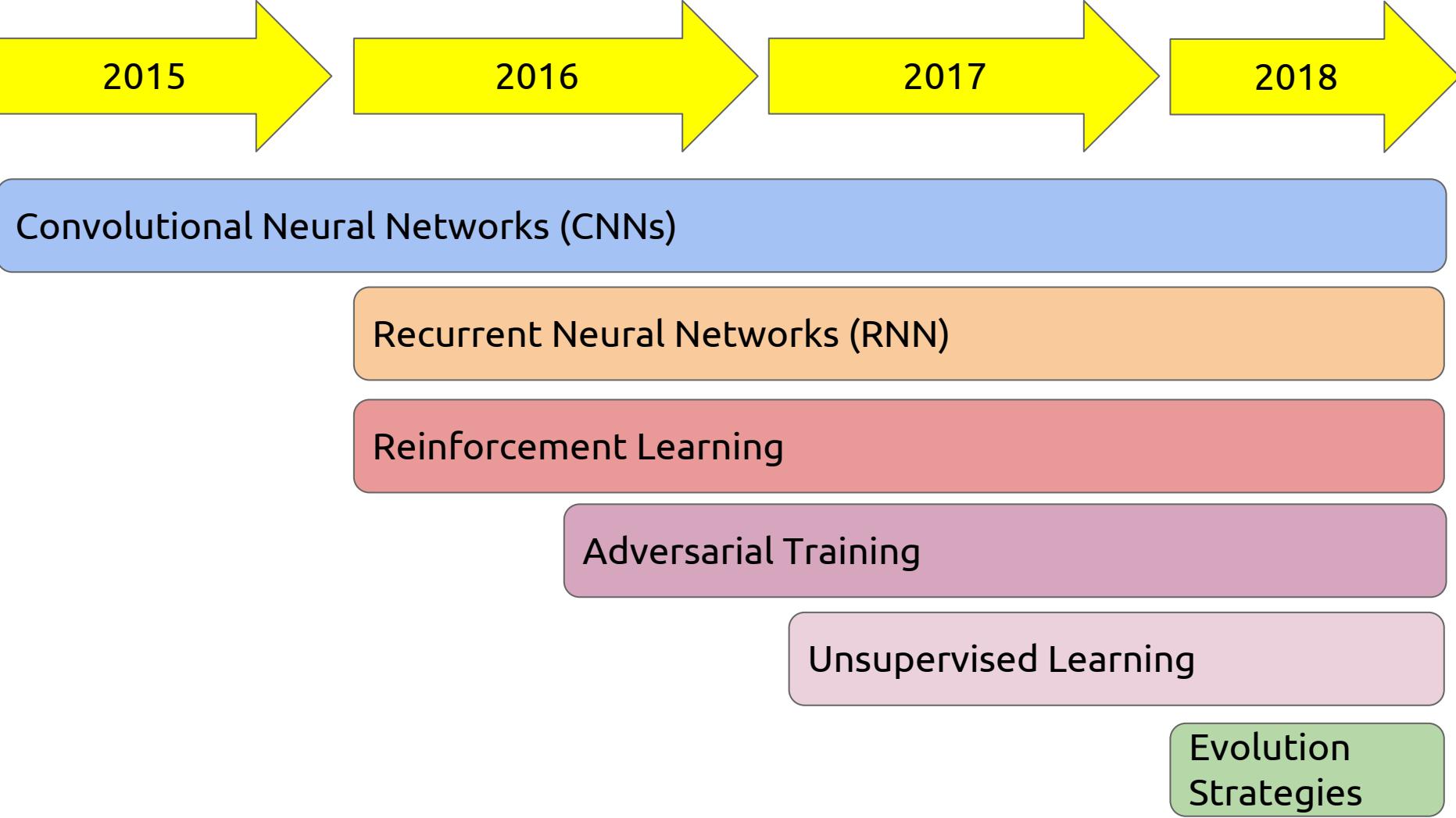
Carlos  
Arenas



Andreu  
Girbau



Eduard  
Ramon



2015

2016

2017

2018

Convolutional Neural Networks (CNNs)

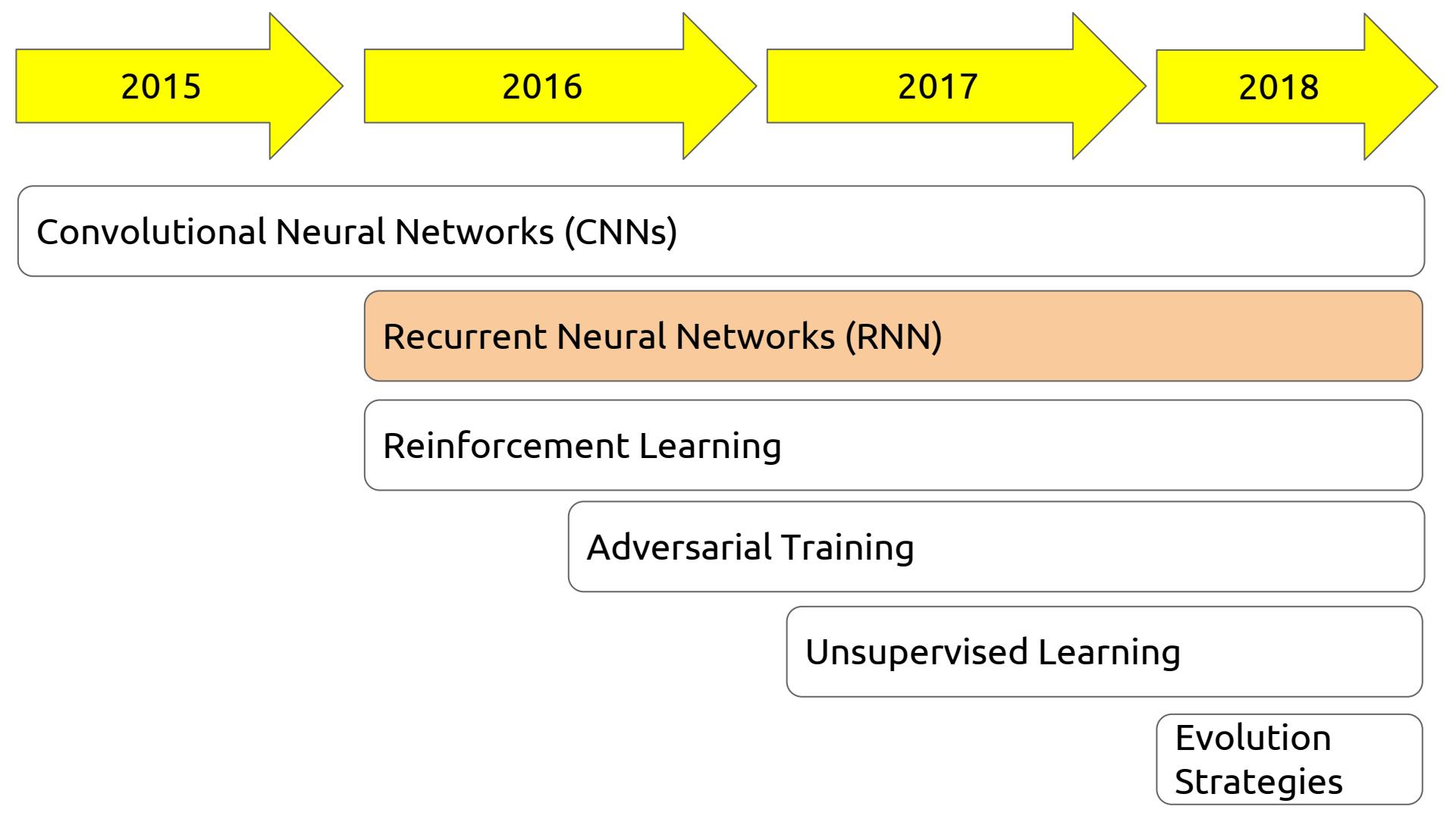
Recurrent Neural Networks (RNN)

Reinforcement Learning

Adversarial Training

Unsupervised Learning

Evolution  
Strategies



2015

2016

2017

2018

Convolutional Neural Networks (CNNs)

Recurrent Neural Networks (RNN)

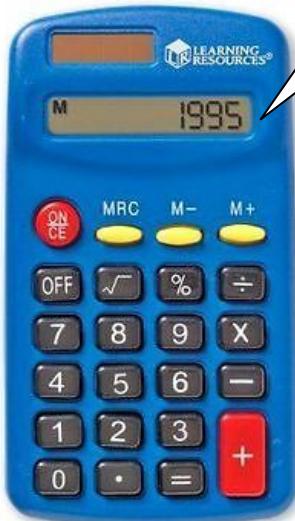
Reinforcement Learning

Adversarial Training

Unsupervised Learning

Evolution  
Strategies

# Motivation: Adaptive Computation Time



$$\int \sqrt{\tan x} dx$$

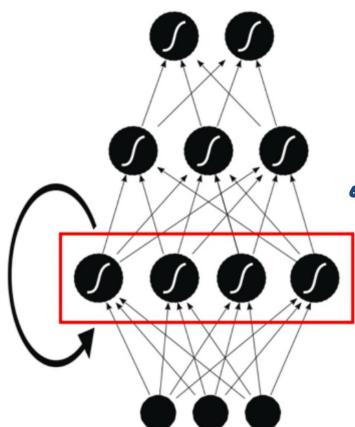
2+2



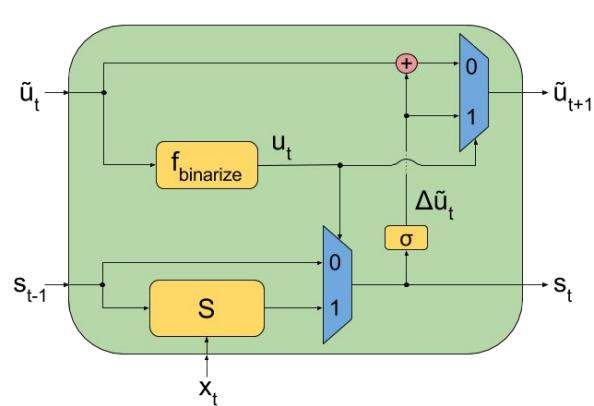


# Outline

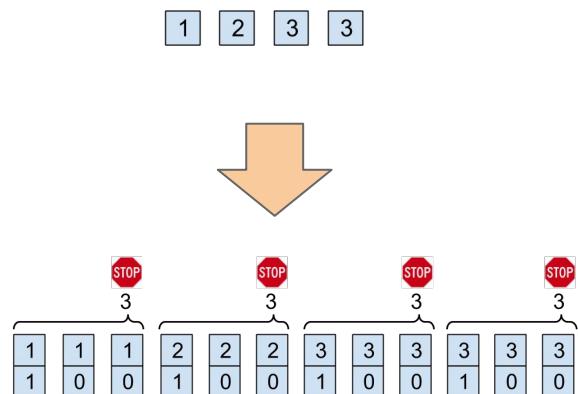
Recurrent Neural Networks (RNNs)



SkipRNN  
[ICLR 2018]



RepeatRNN  
[ICLRW 2018]

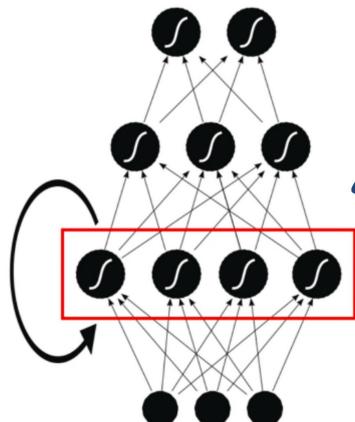


# Outline

Recurrent Neural  
Networks (RNNs)

SkipRNN  
[ICLR 2018]

RepeatRNN  
[ICLRW 2018]





# Feed-forward Neural Network

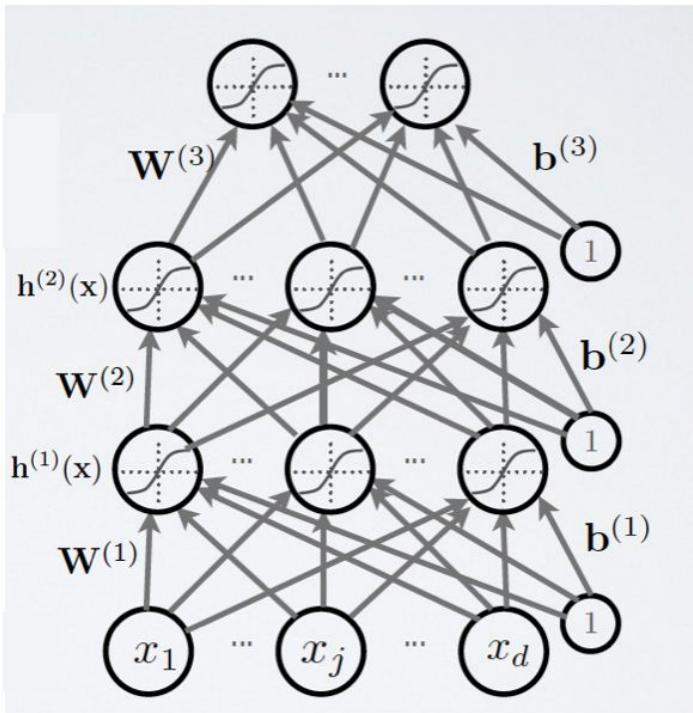
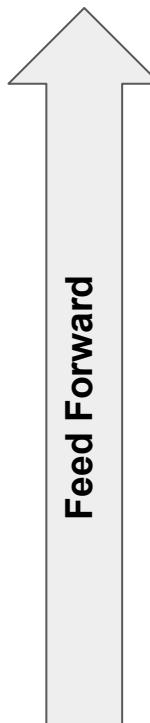


Figure: Hugo Larochelle

Hidden States  $h_1$  &  $h_2$

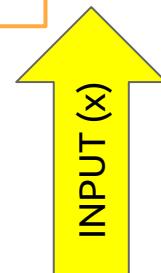


Feed-forward Weights ( $W_i$ )

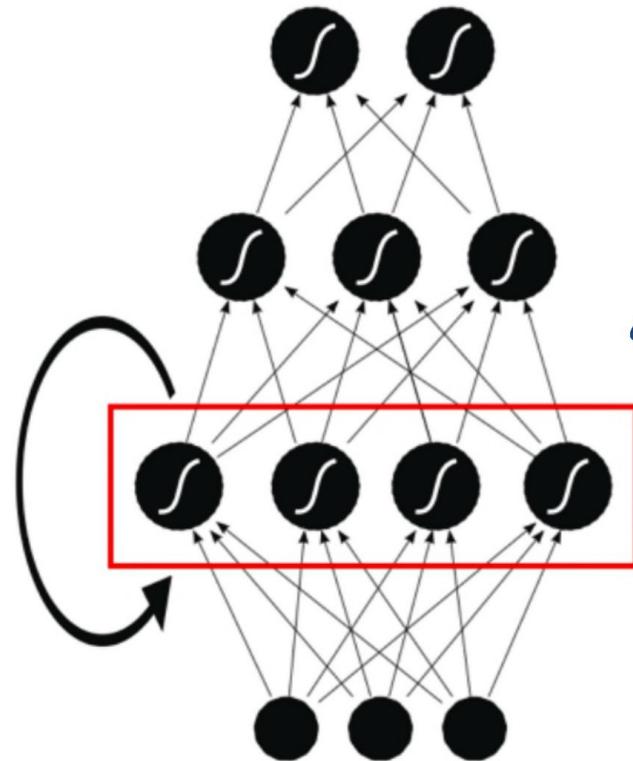
$$y = f(W_3 \cdot h_2 + b_3)$$

$$h_2 = f(W_2 \cdot h_1 + b_2)$$

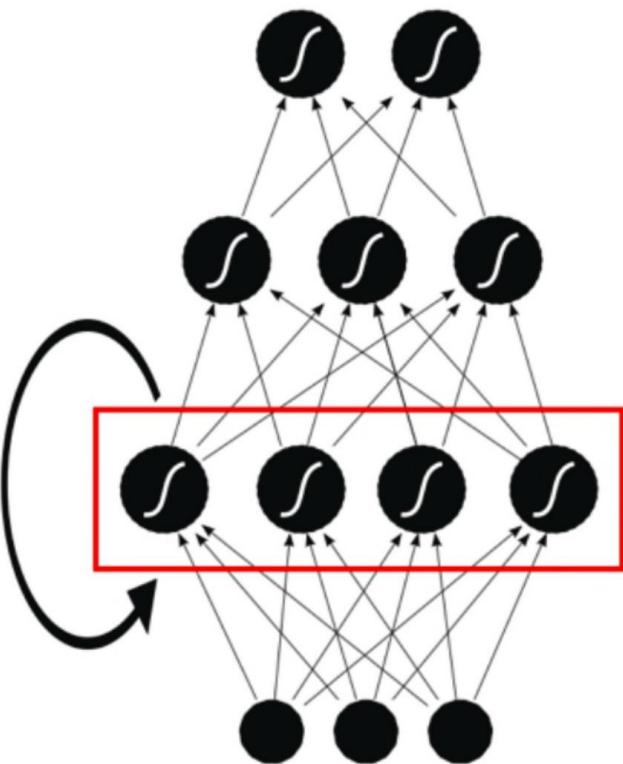
$$h_1 = f(W_1 \cdot x + b_1)$$



# Recurrent Neural Network

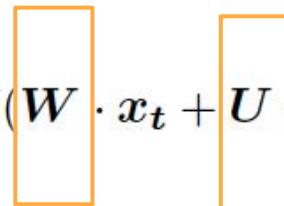


# Recurrent Neural Network



Feed-forward  
Weights ( $W$ )

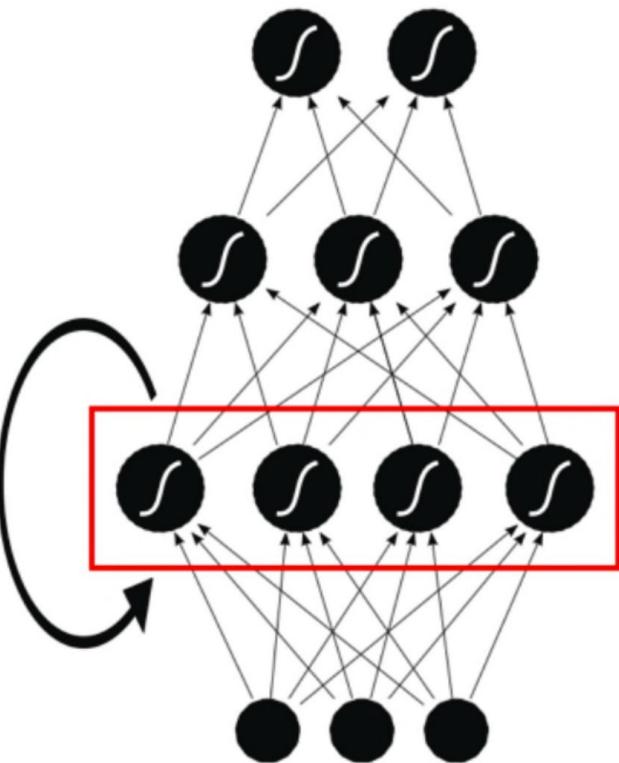
$$h_t = f(W \cdot x_t + U \cdot h_{t-1} + b)$$



Recurrent  
Weights ( $U$ )



# Recurrent Neural Network



Updated state

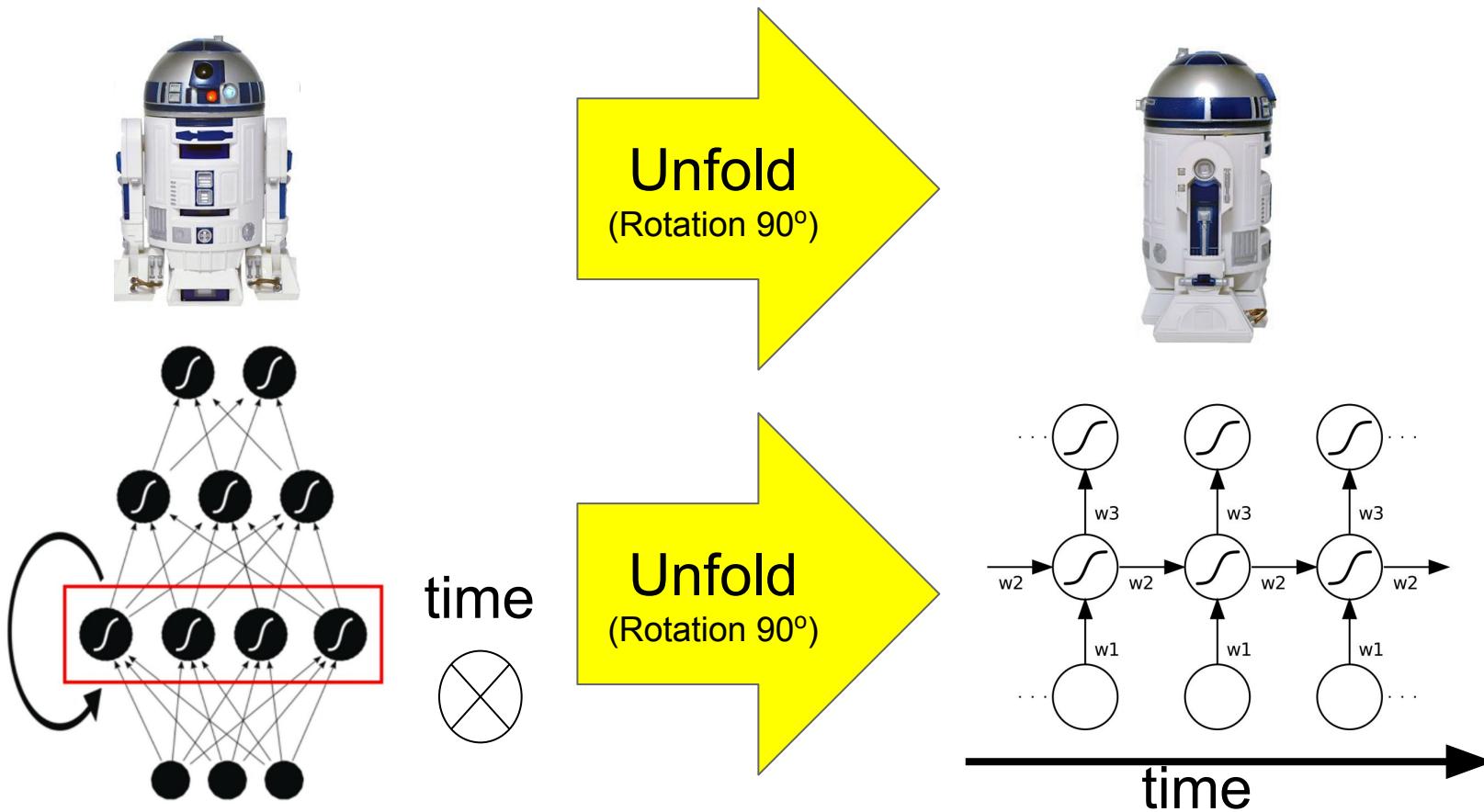
$$h_t = f(W \cdot x_t + U \cdot h_{t-1} + b)$$

INPUT ( $x$ )

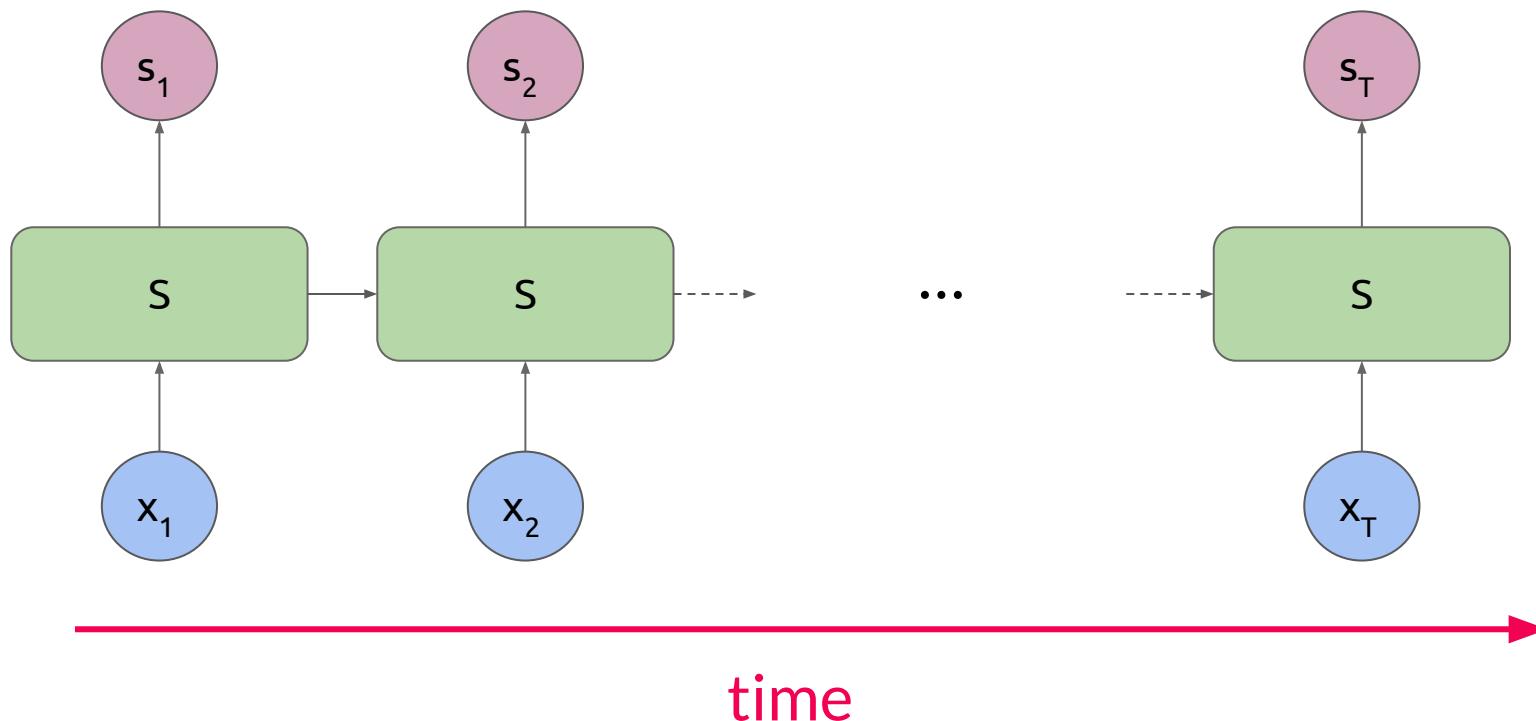
Previous state

$$h_{t-1}$$

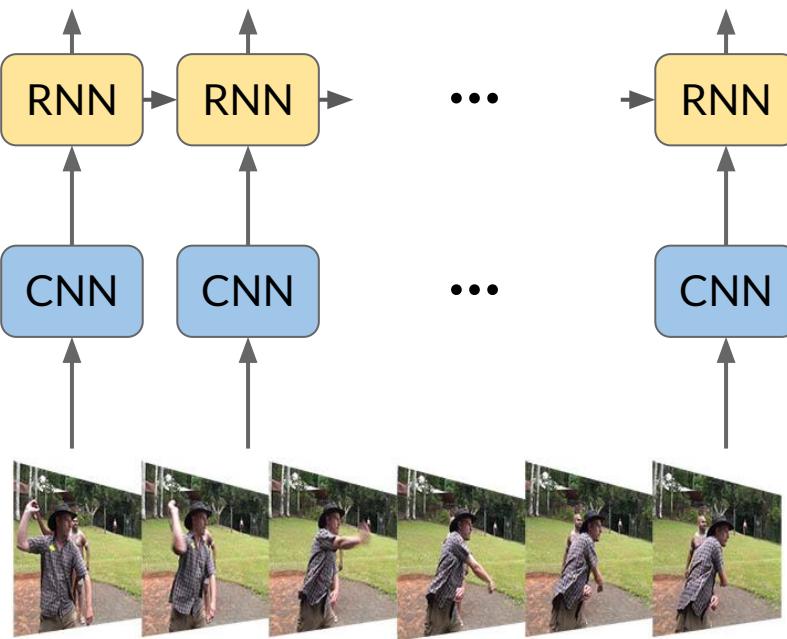
# Recurrent Neural Network



# Recurrent Neural Network

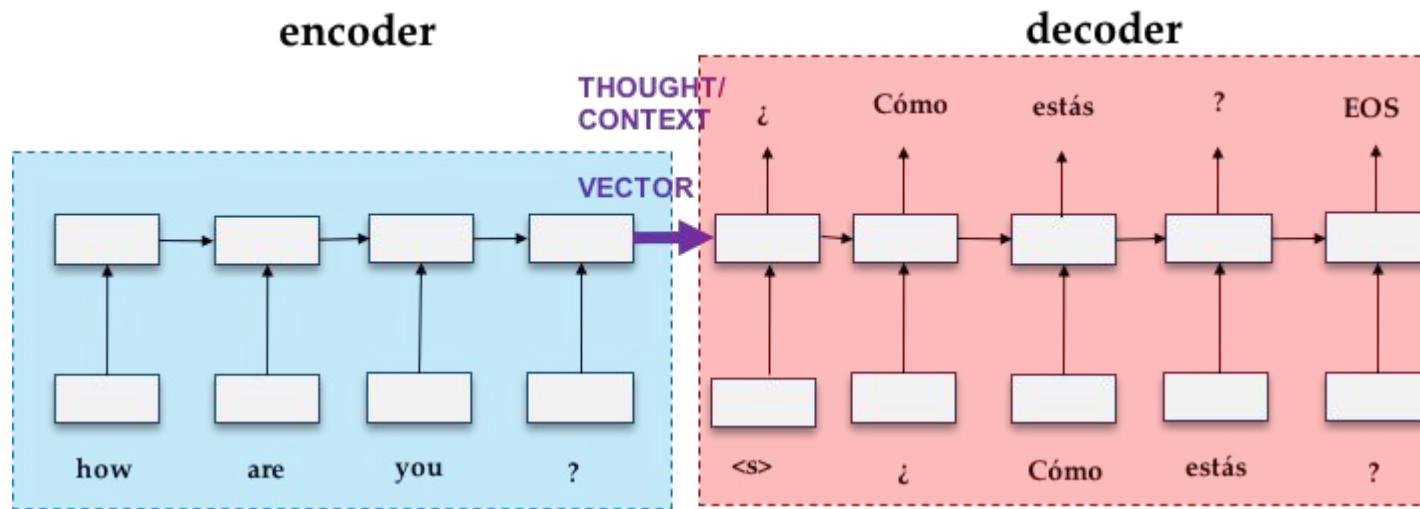


# Recurrent Neural Network



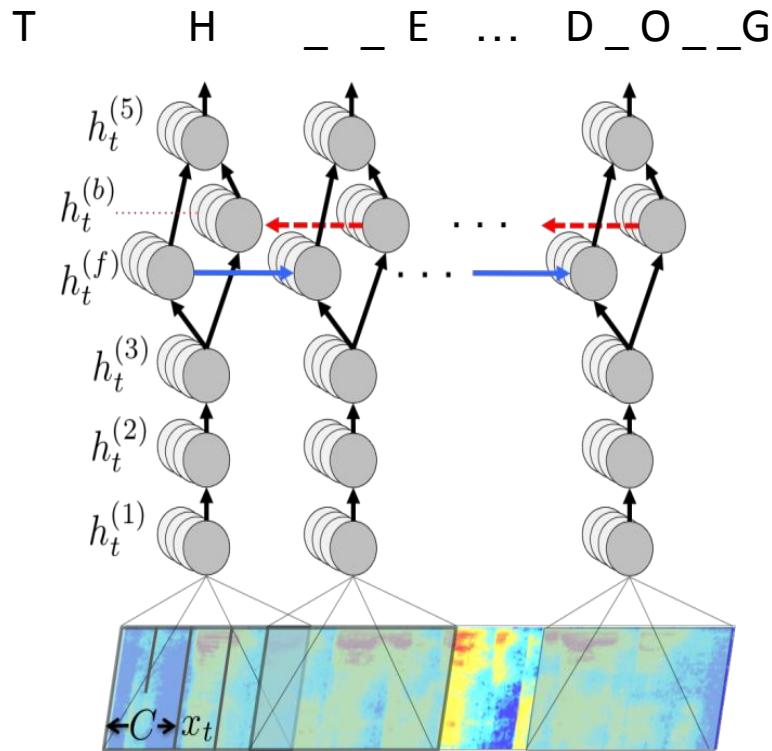
Video sequences (eg. action recognition)

# Recurrent Neural Network



Word sequences (eg. Machine Translation)

# Recurrent Neural Network



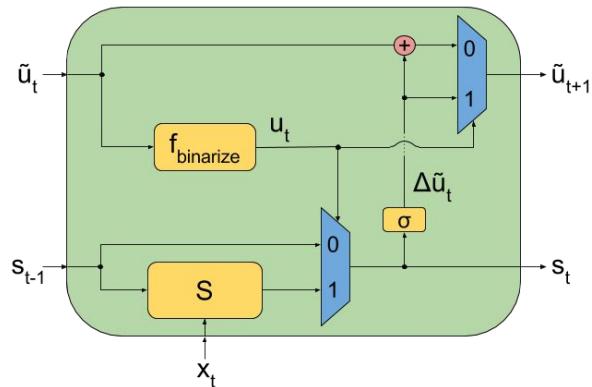
Spectrograms sequence (eg. speech recognition)

# Outline

Recurrent Neural Networks (RNNs)

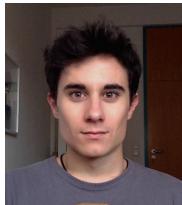
SkipRNN  
[ICLR 2018]

RepeatRNN  
[ICLRW 2018]





# Skip RNN: Learning to Skip State Updates in RNNs



[Víctor Campos](#)



[Brendan Jou](#)



[Jordi Torres](#)



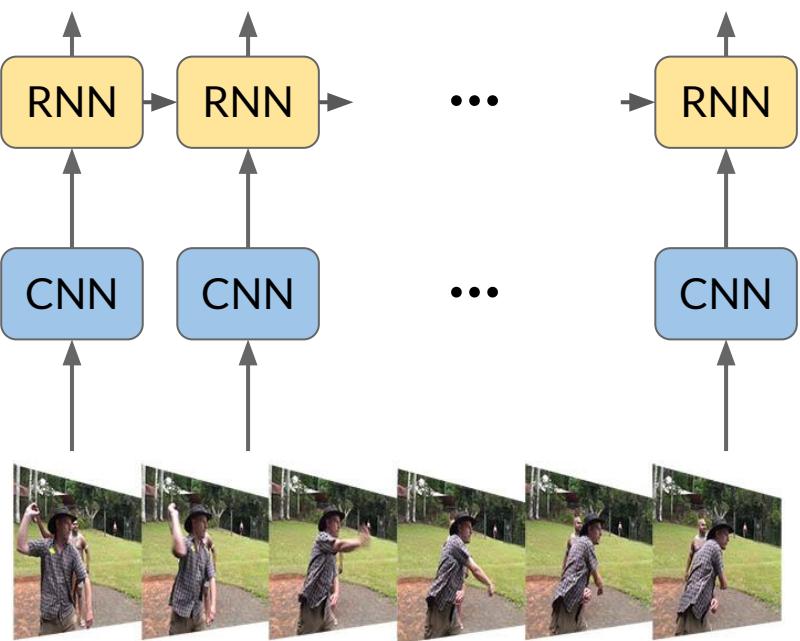
[Xavier Giró-i-Nieto](#)



[Shih-Fu Chang](#)

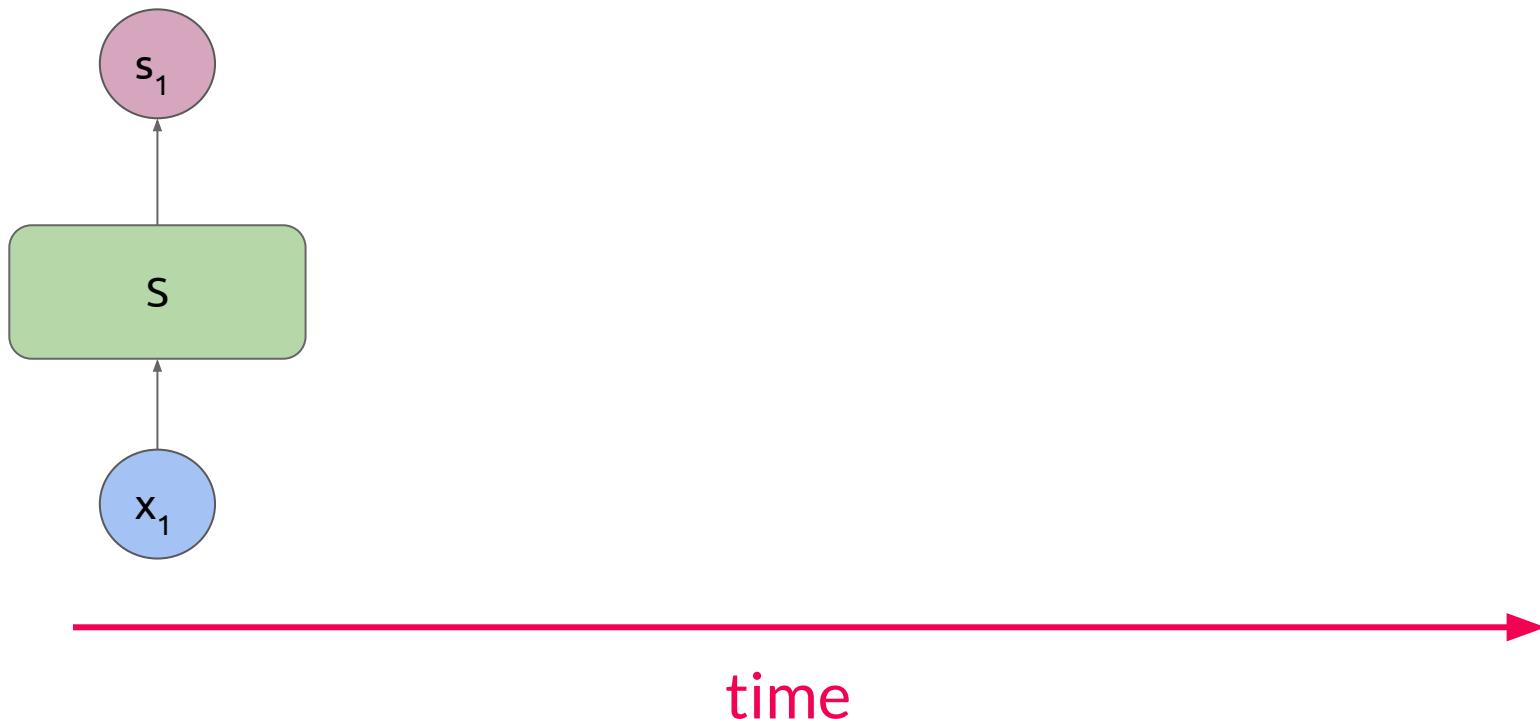
Victor Campos, Brendan Jou, Xavier Giro-i-Nieto, Jordi Torres, and Shih-Fu Chang. ["Skip RNN: Learning to Skip State Updates in Recurrent Neural Networks"](#), ICLR 2018.

# Motivation

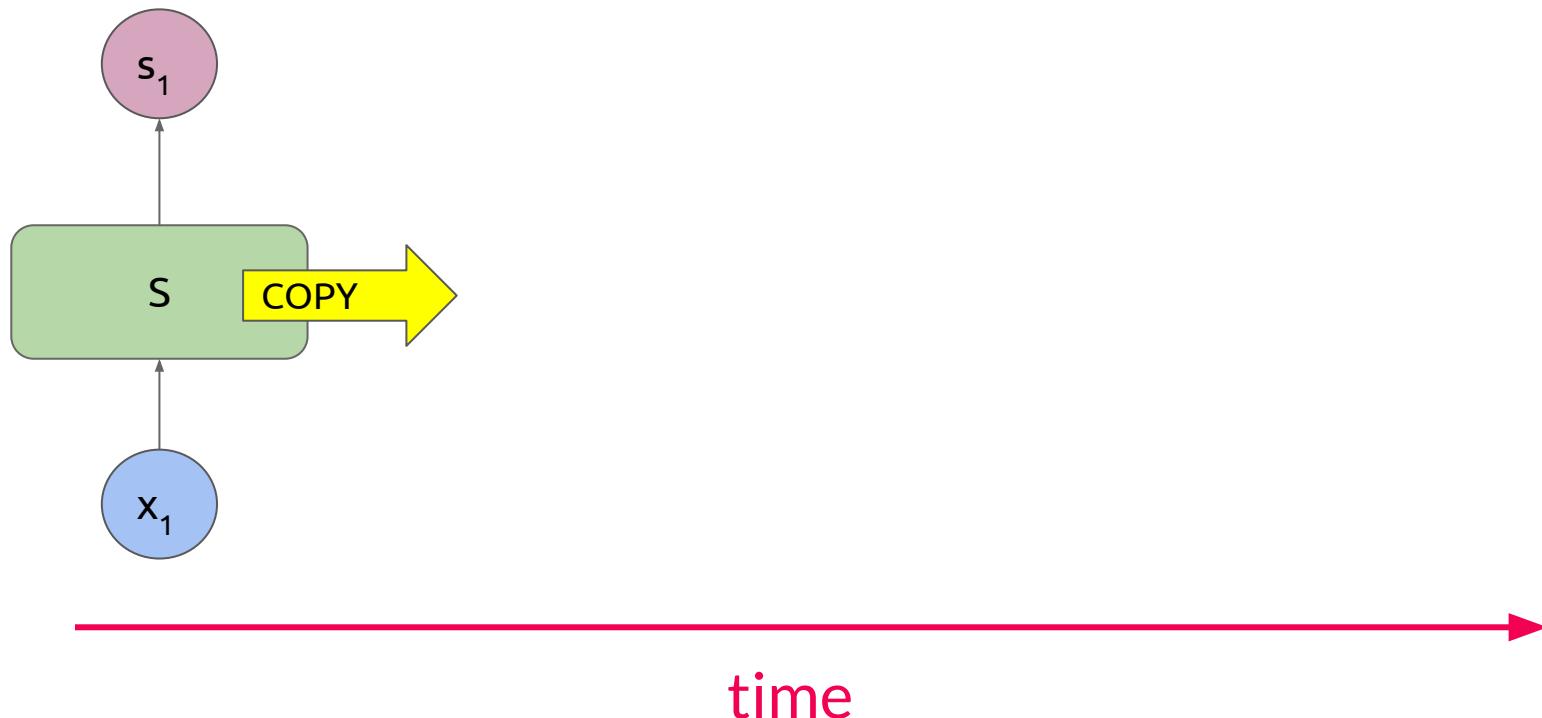


Used      Unused

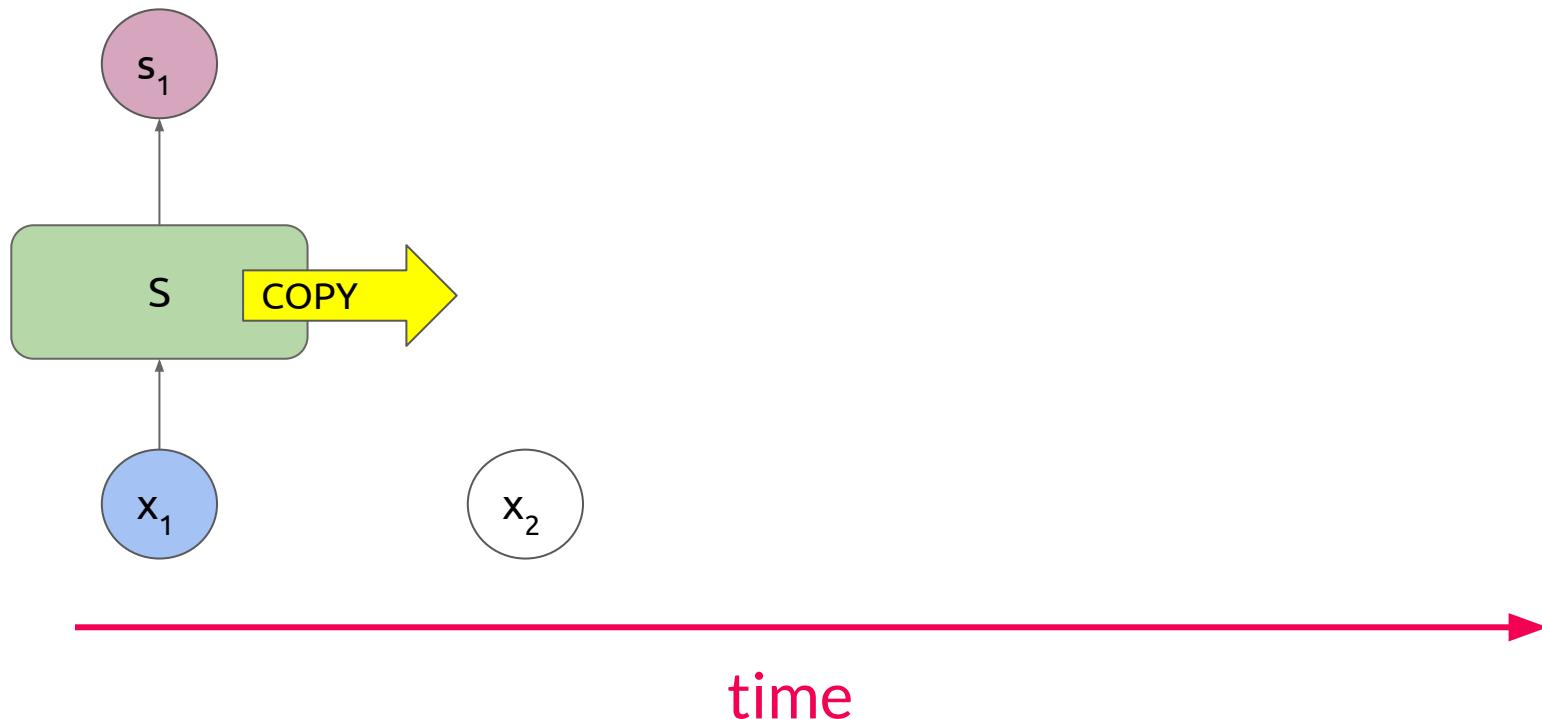
# SkipRNN



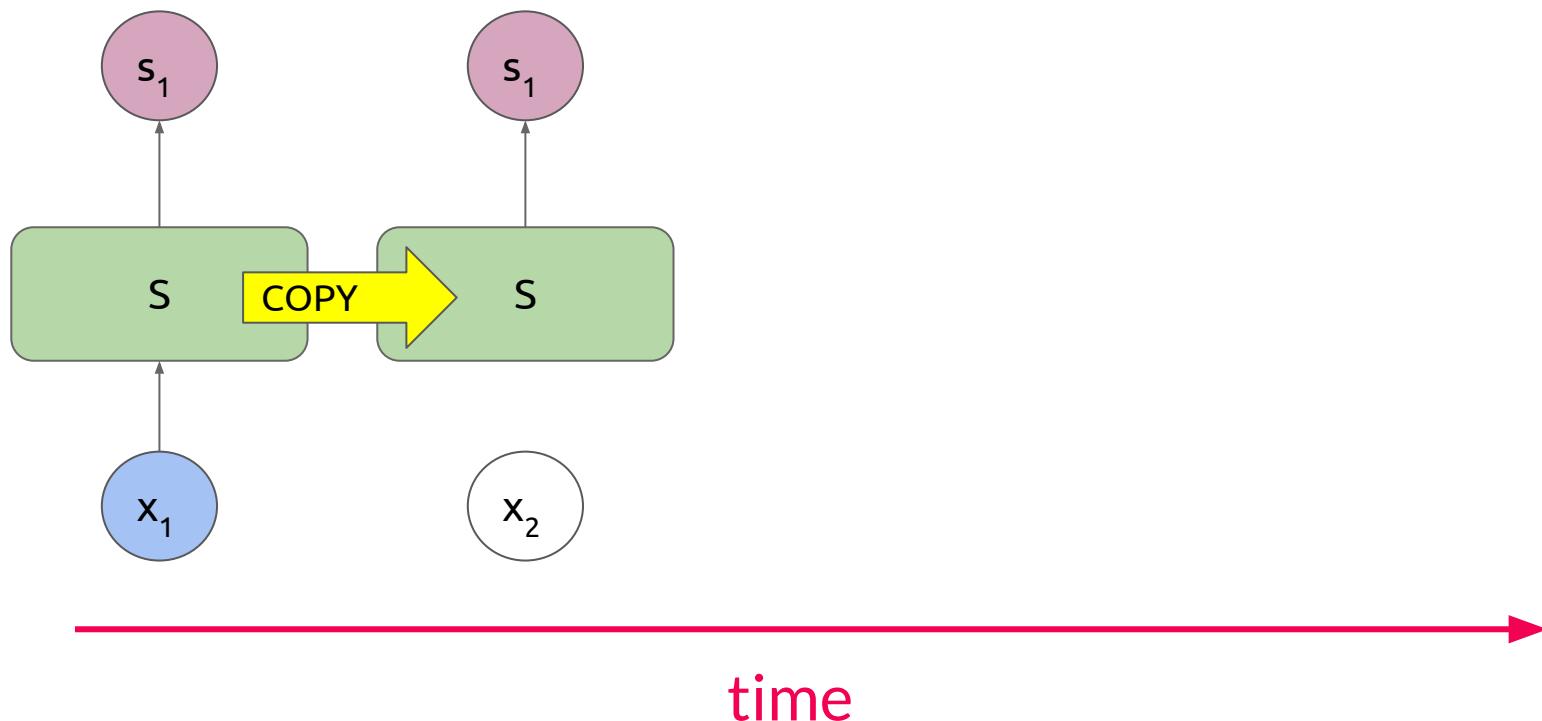
# SkipRNN



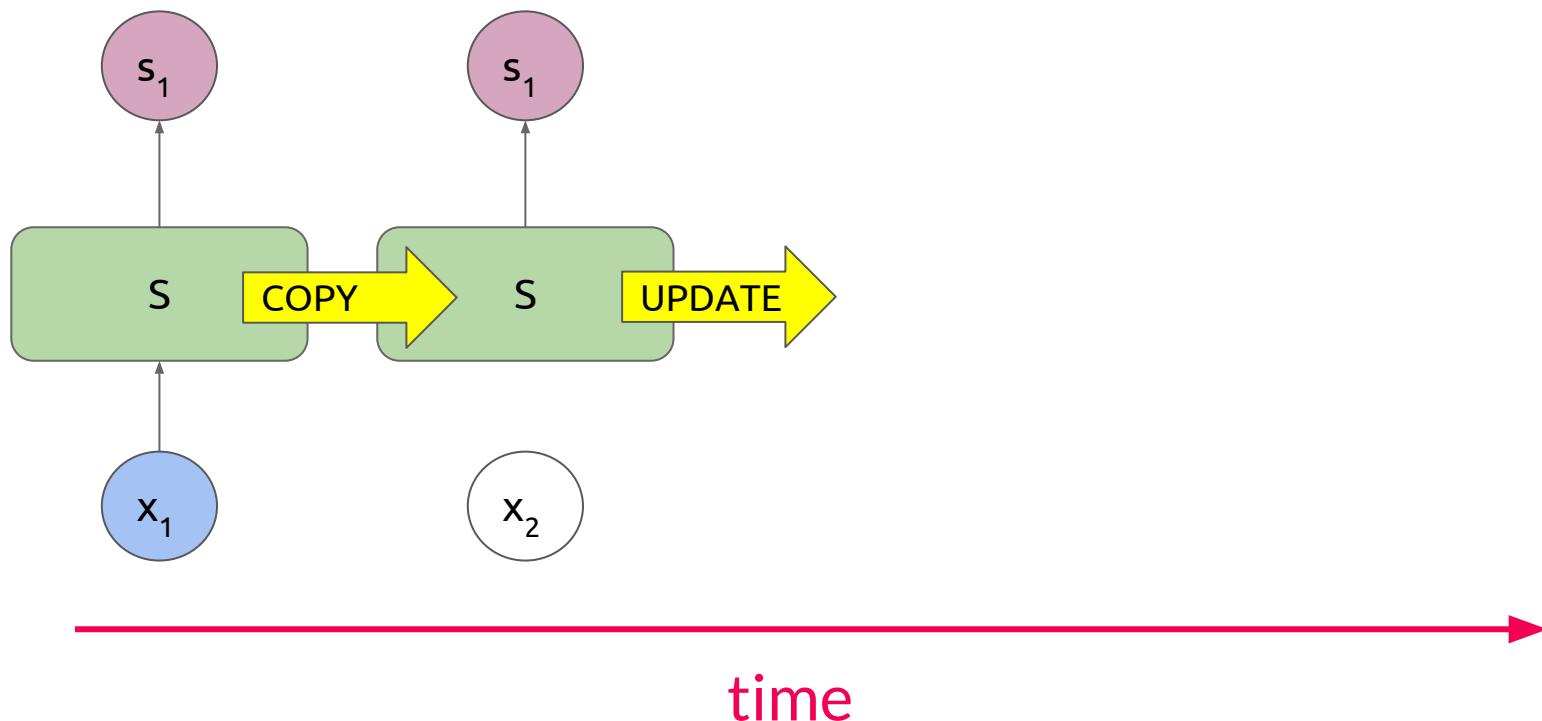
# SkipRNN



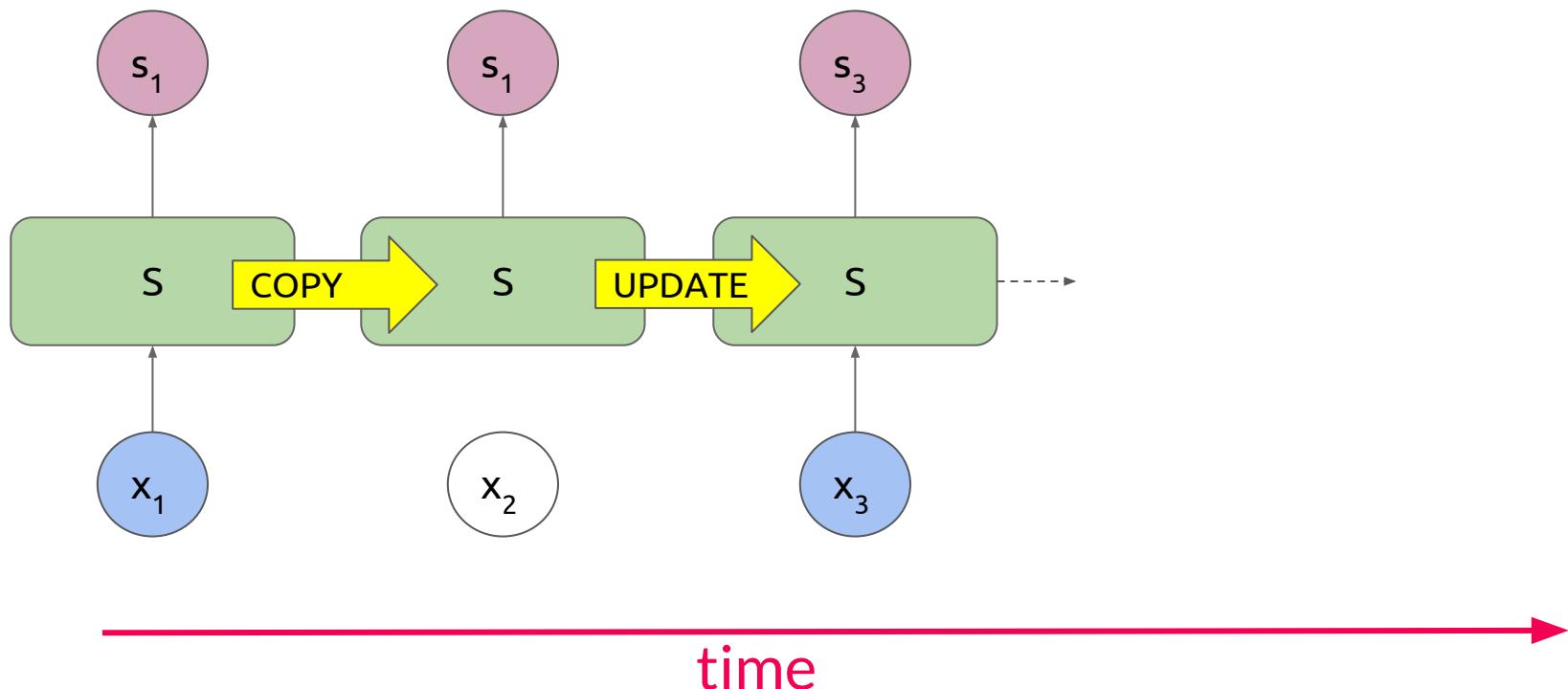
# SkipRNN



# SkipRNN

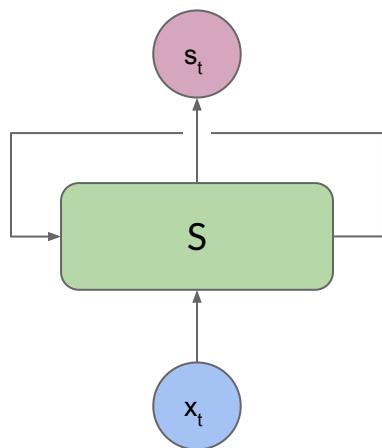


# SkipRNN



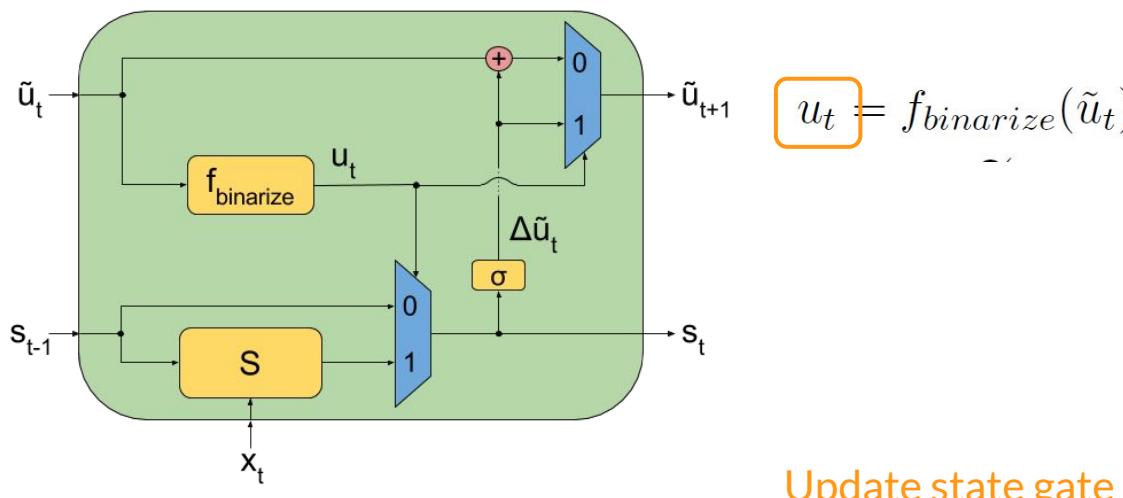
# SkipRNN

Intuition: introduce a binary *update state gate*,  $u_t$ , deciding whether the RNN state is updated or copied



$$s_t = \begin{cases} S(x_t, h_{t-1}) & \text{if } u_t = 1 \quad // \text{update operation} \\ s_{t-1} & \text{if } u_t = 0 \quad // \text{copy operation} \end{cases}$$

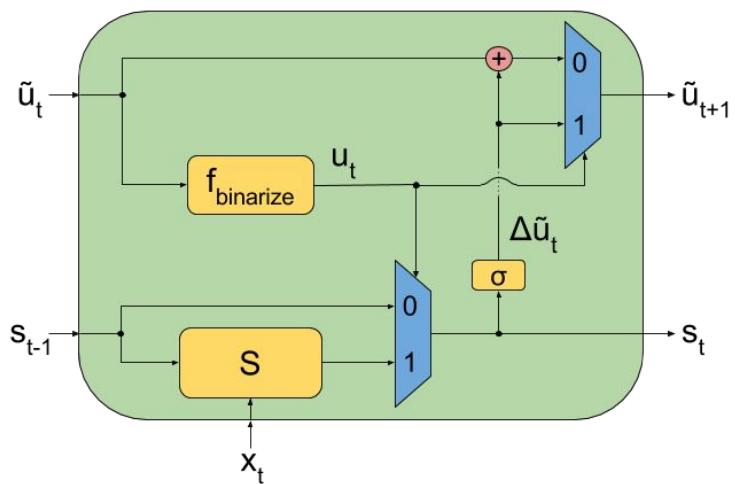
# SkipRNN



$$u_t = f_{\text{binarize}}(\tilde{u}_t)$$

Update state gate  $\in \{0, 1\}$

# SkipRNN



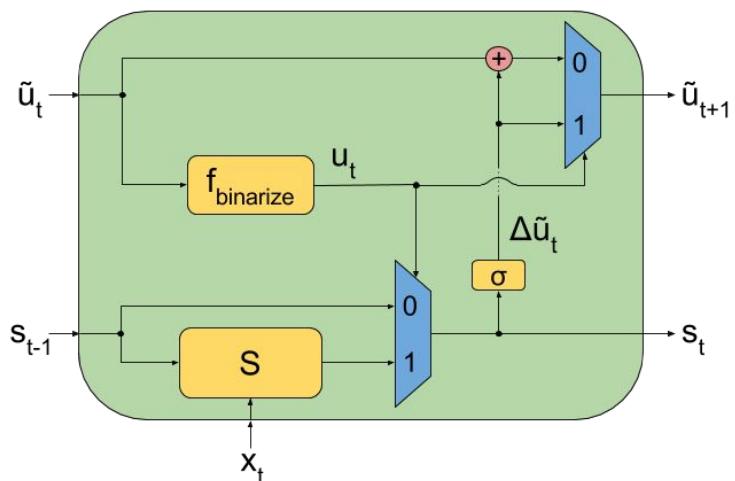
$$u_t = f_{\text{binarize}}(\tilde{u}_t)$$

$$\tilde{u}_{t+1} = u_t \cdot \Delta \tilde{u}_t + (1 - u_t) \cdot (\tilde{u}_t + \min(\Delta \tilde{u}_t, 1 - \tilde{u}_t))$$

Update state gate  $\in \{0, 1\}$

Update state probability  $\in [0, 1]$

# SkipRNN



$$u_t = f_{\text{binarize}}(\tilde{u}_t)$$

$$s_t = u_t \cdot S(s_{t-1}, x_t) + (1 - u_t) \cdot s_{t-1}$$

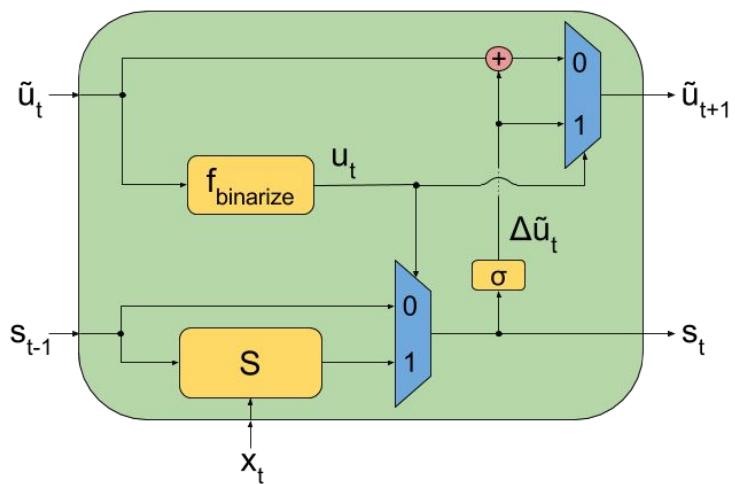
$$\Delta \tilde{u}_t = \sigma(W_p s_t + b_p)$$

$$\tilde{u}_{t+1} = u_t \cdot \Delta \tilde{u}_t + (1 - u_t) \cdot (\tilde{u}_t + \min(\Delta \tilde{u}_t, 1 - \tilde{u}_t))$$

Update state gate  $\in \{0, 1\}$

Update state probability  $\in [0, 1]$

# SkipRNN



$$u_t = f_{\text{binarize}}(\tilde{u}_t)$$

$$s_t = u_t \cdot S(s_{t-1}, x_t) + (1 - u_t) \cdot s_{t-1}$$

$$\Delta \tilde{u}_t = \sigma(W_p s_t + b_p)$$

$$\tilde{u}_{t+1} = u_t \cdot \Delta \tilde{u}_t + (1 - u_t) \cdot (\tilde{u}_t + \min(\Delta \tilde{u}_t, 1 - \tilde{u}_t))$$

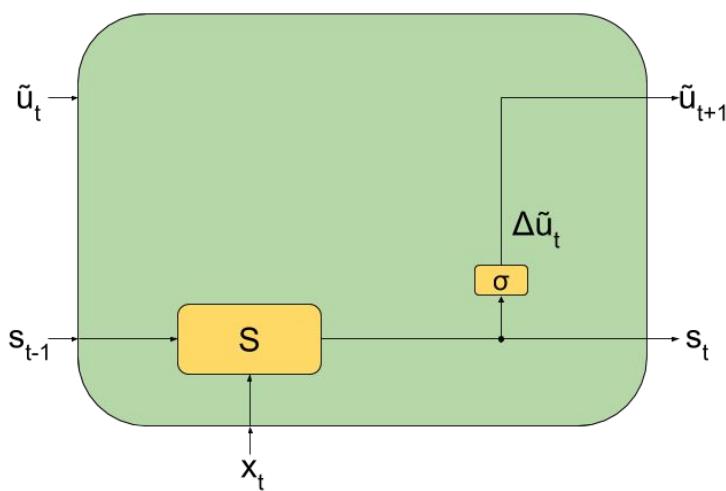
Update state gate  $\in \{0, 1\}$

Update state probability  $\in [0, 1]$

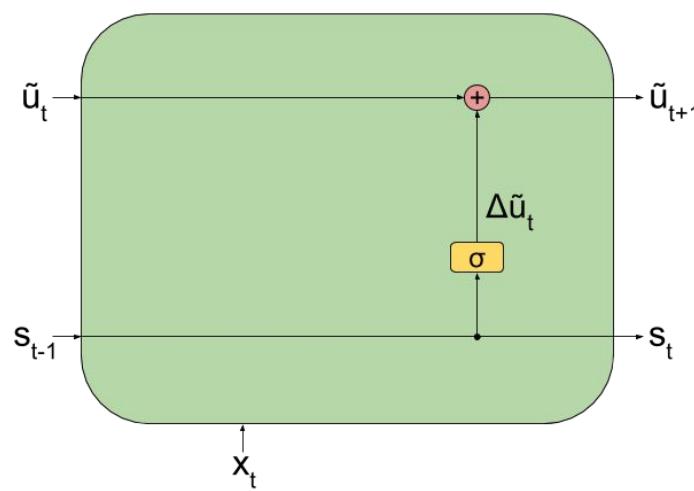
Increment for the update state probability

# SkipRNN

Update state ( $u_t = 1$ )

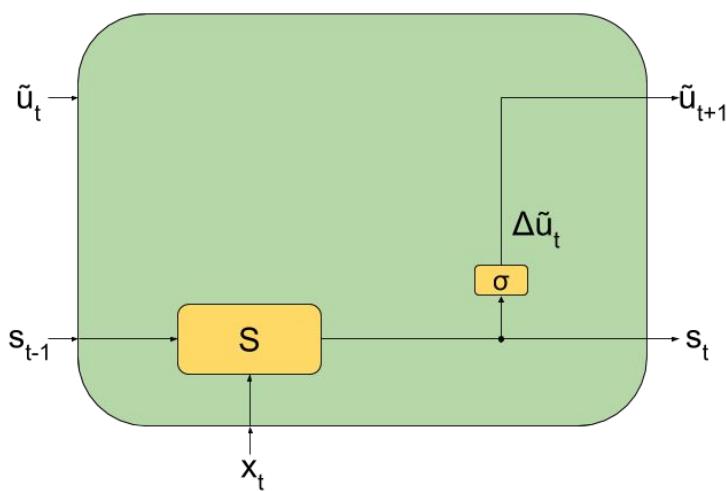


Copy state ( $u_t = 0$ )

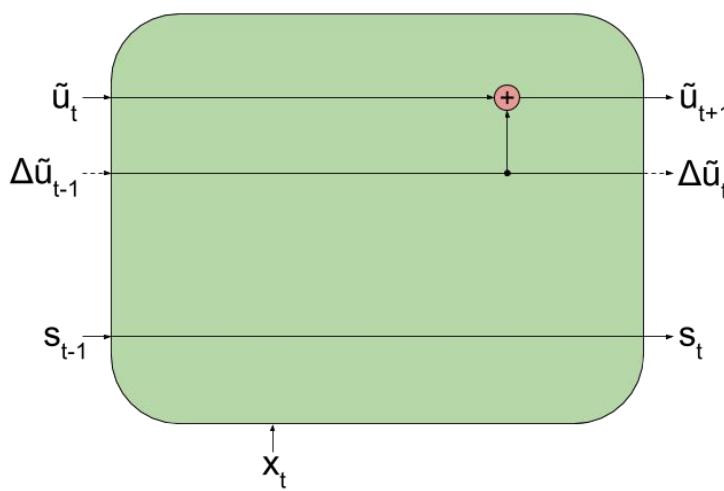


# SkipRNN

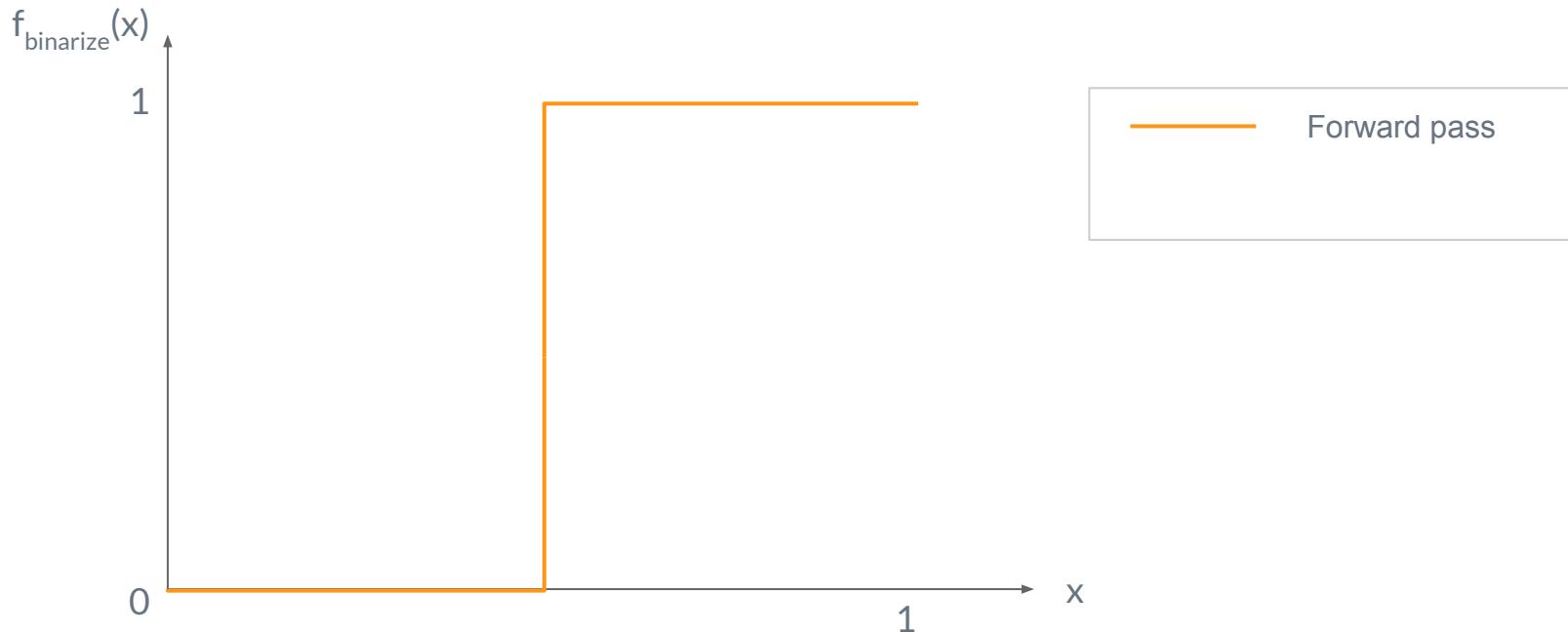
Update state ( $u_t = 1$ )



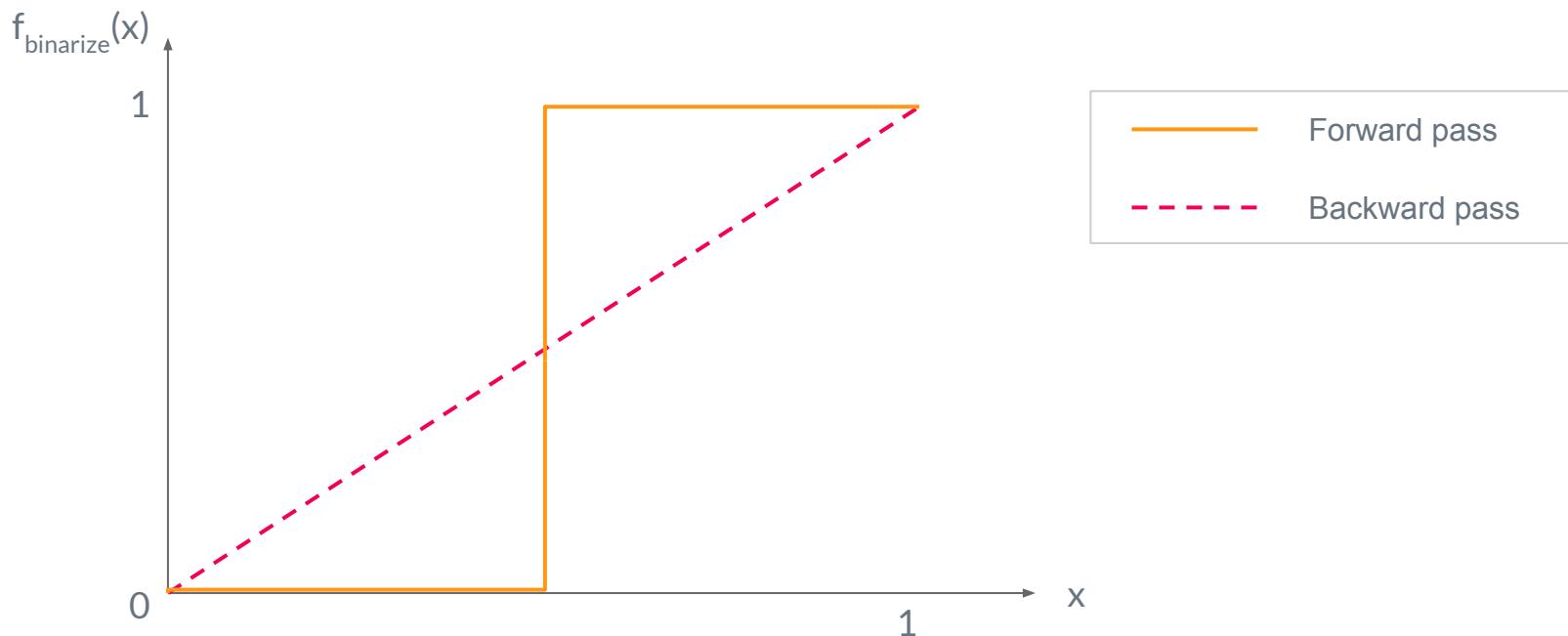
Copy state ( $u_t = 0$ )



# Straight Through Estimator



# Straight Through Estimator



# Limiting computation

Intuition: the network can be encouraged to perform fewer updates by adding a penalization when  $u_t = 1$



$$L_{budget} = \lambda \cdot \sum_{t=1}^T u_t \rightarrow \begin{cases} 1 & \text{if sample used} \\ 0 & \text{otherwise} \end{cases}$$

cost per sample

# Experiments: Sequential MNIST

Used  
Unused

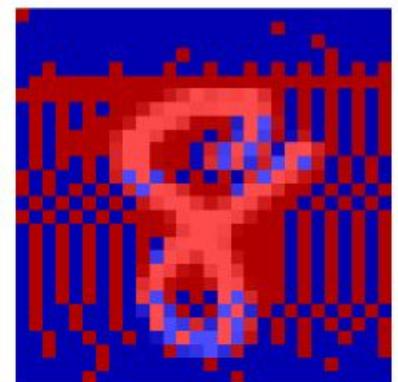
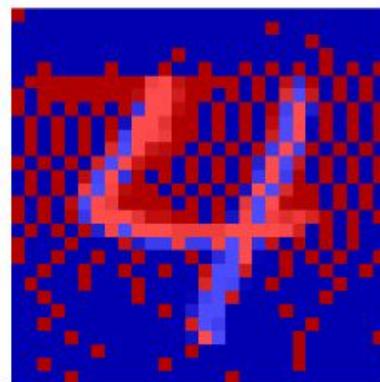
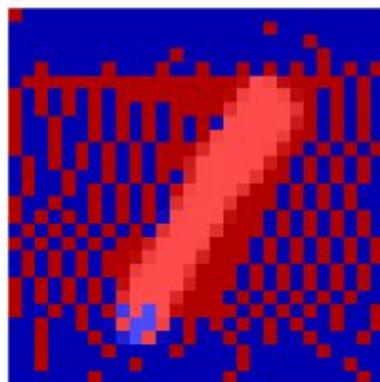
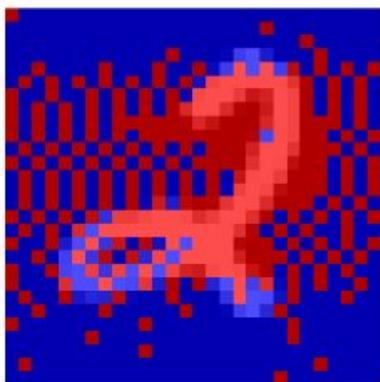


Figure 3: Sample usage examples for the Skip LSTM with  $\lambda = 10^{-4}$  on the test set of MNIST. Red pixels are used, whereas blue ones are skipped.



# Experiments: Sequential MNIST

Model	Accuracy	State updates	Inference FLOPs
LSTM	$0.910 \pm 0.045$	$784.00 \pm 0.00$	$3.83 \times 10^7$
LSTM ( $p_{skip} = 0.5$ )	$0.893 \pm 0.003$	$392.03 \pm 0.05$	$1.91 \times 10^7$
Skip LSTM, $\lambda = 10^{-4}$	$0.973 \pm 0.002$	$379.38 \pm 33.09$	$1.86 \times 10^7$

Better  
accuracy...

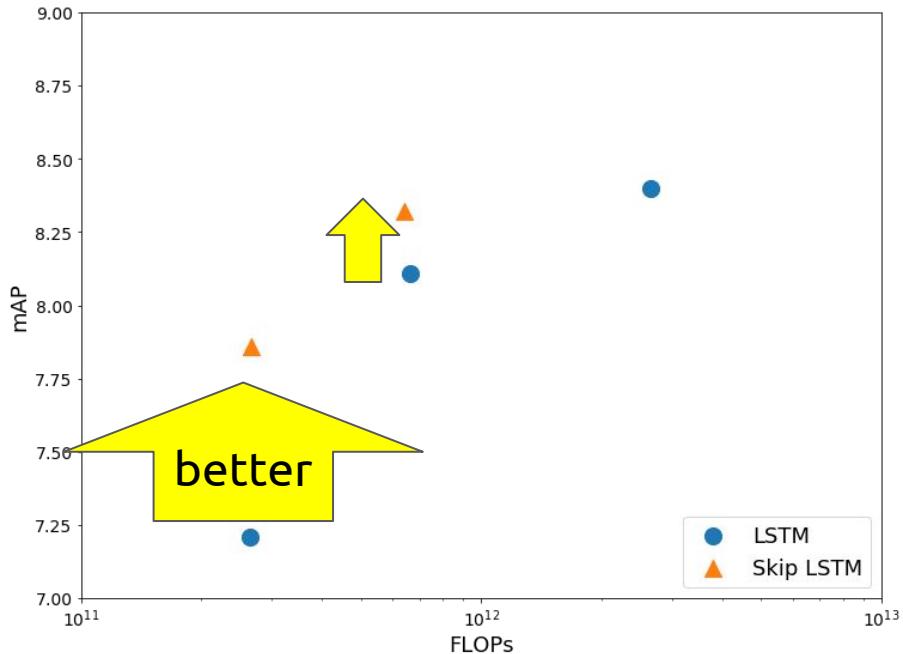
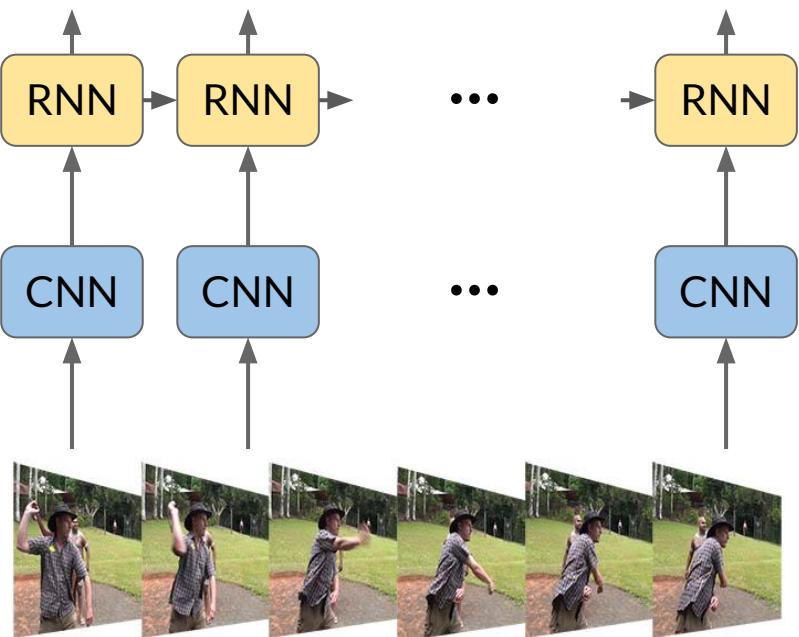
# Experiments: Sequential MNIST

Model	Accuracy	State updates	Inference FLOPs
LSTM	$0.910 \pm 0.045$	$784.00 \pm 0.00$	$3.83 \times 10^7$
LSTM ( $p_{skip} = 0.5$ )	$0.893 \pm 0.003$	$392.03 \pm 0.05$	$1.91 \times 10^7$
Skip LSTM, $\lambda = 10^{-4}$	$0.973 \pm 0.002$	$379.38 \pm 33.09$	$1.86 \times 10^7$

...with less  
computation

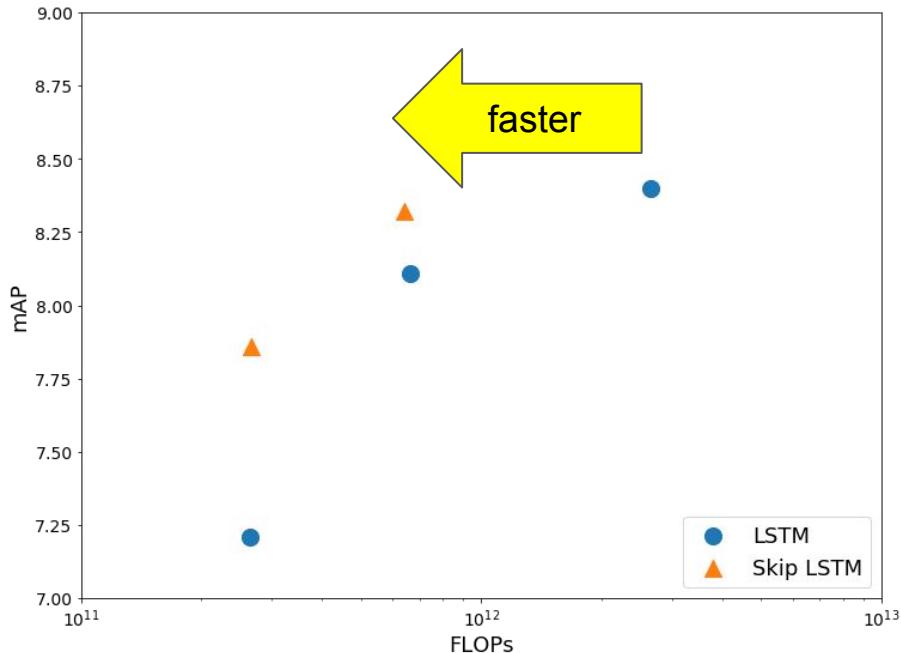
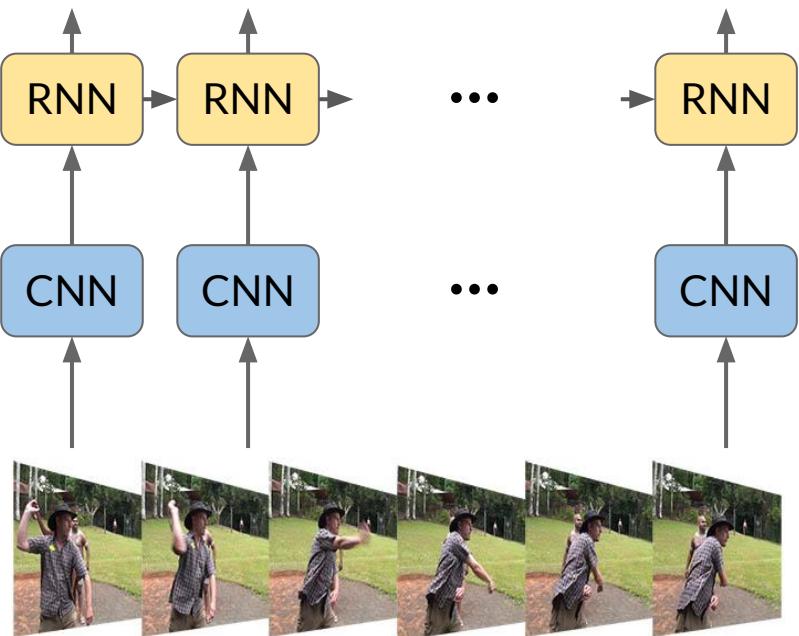


# Experiments: Action Localization





# Experiments: Action Localization





# Open science

AUTHORS INTRODUCTION MODEL RESULTS EXAMPLES CODE ACKNOWLEDGEMENTS

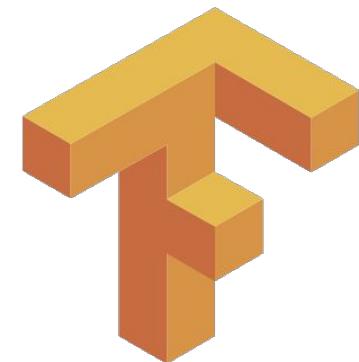
Fork me on GitHub

## Skip RNN: Skipping State Updates in Recurrent Neural Networks

Victor Campos      Brendan Jou      Jordi Torres      Xavier Giró-i Nieto      Shih-Fu Chang



<https://git.io/skip-rnn>



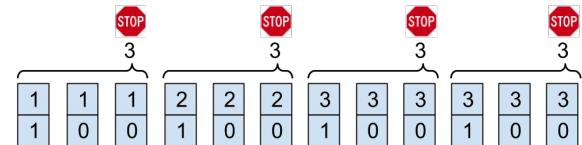
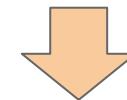
# Outline

Recurrent Neural  
Networks (RNNs)

SkipRNN  
[ICLR 2018]

RepeatRNN  
[ICLRW 2018]

1 2 3 3



# Reproducing and Analyzing Adaptive Computation Time in PyTorch and TensorFlow



Dani Fojo



Víctor Campos

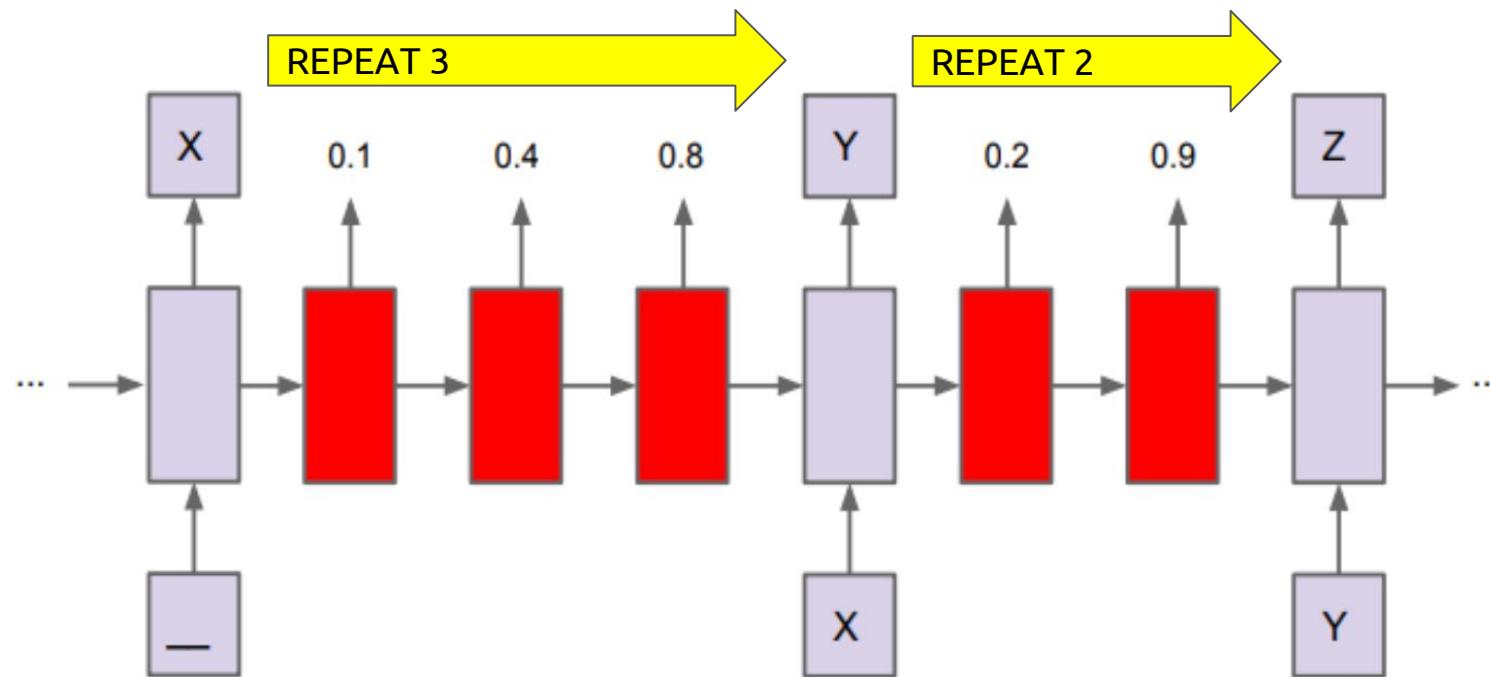


Xavier Giró-i-Nieto

Fojo, Daniel, Víctor Campos, and Xavier Giró-i-Nieto. "[Comparing Fixed and Adaptive Computation Time for Recurrent Neural Networks.](#)" ICLR Workshop 2018.

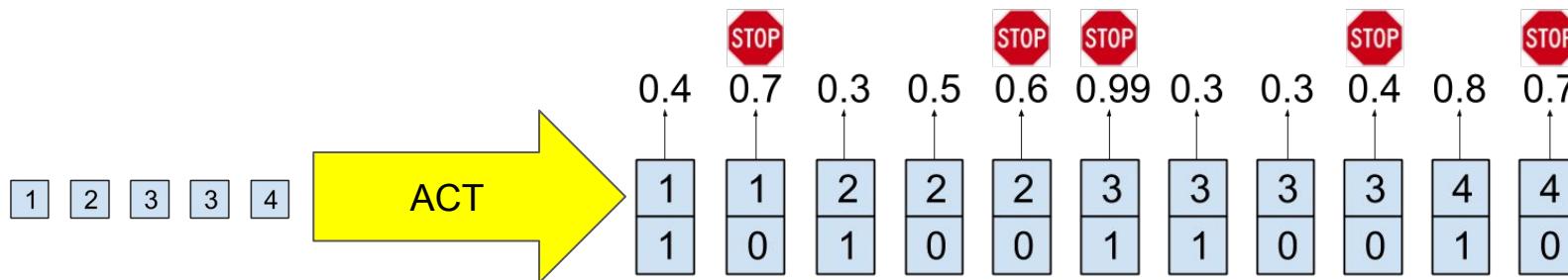


# Adaptive Computation Time (ACT)

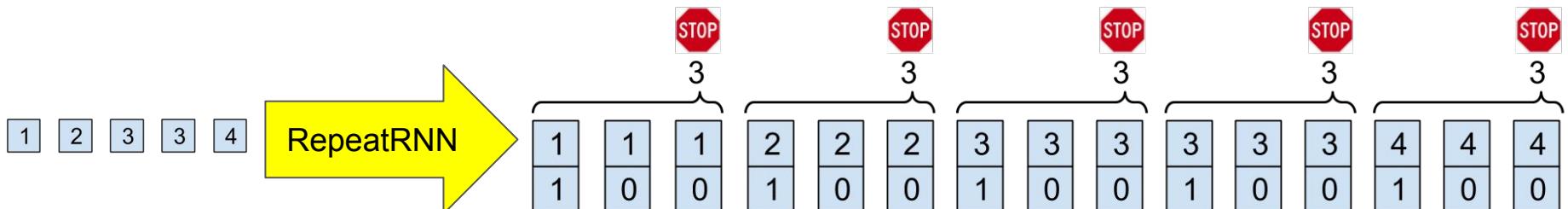
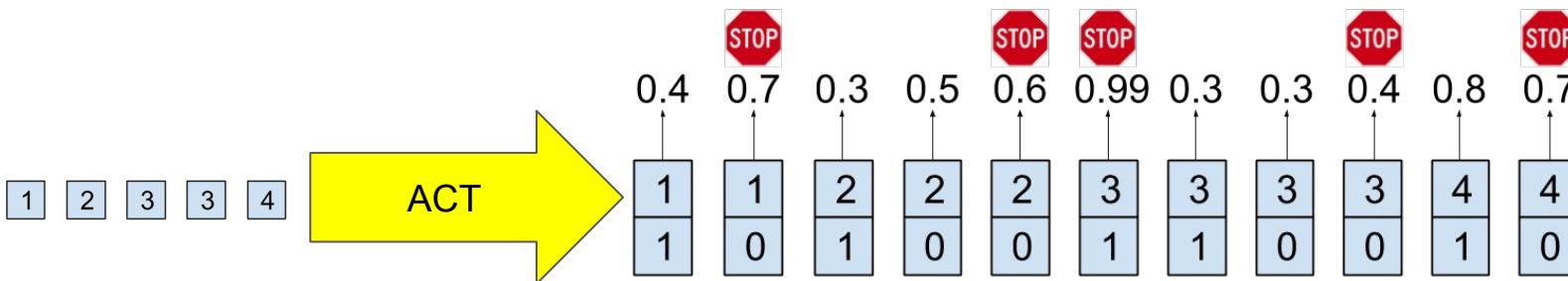


Graves, Alex. "[Adaptive computation time for recurrent neural networks.](#)" arXiv preprint arXiv:1603.08983 (2016).

# Adaptive Computation Time (ACT)

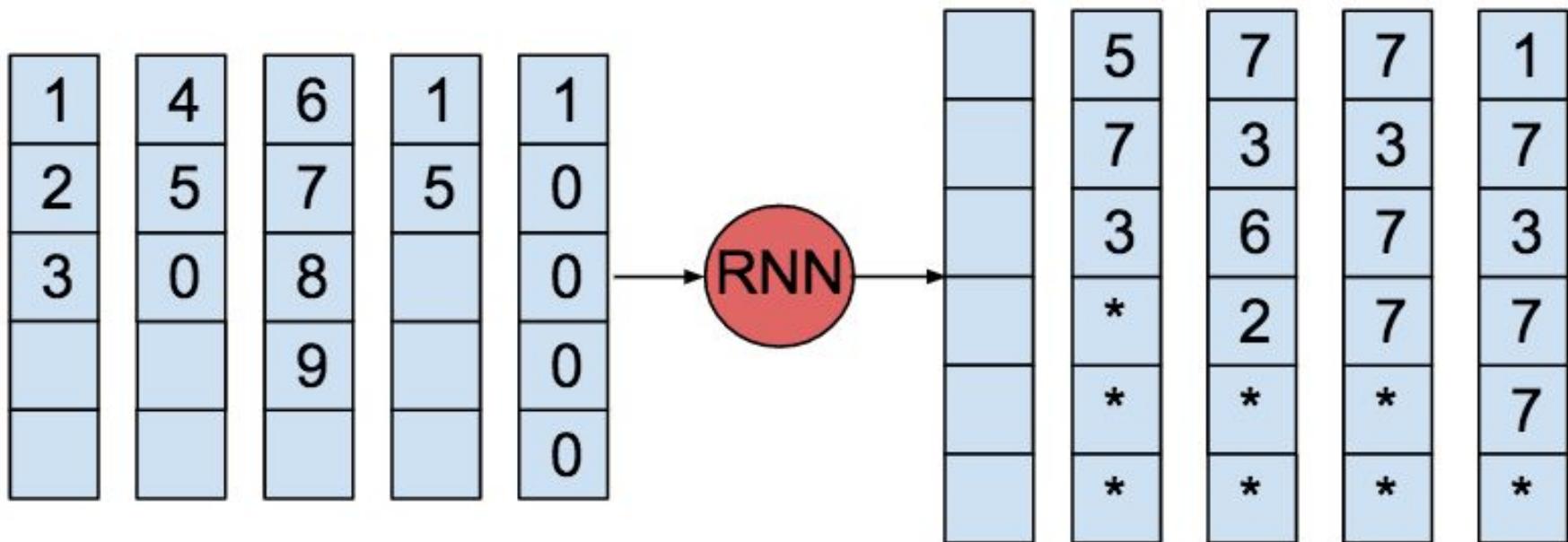


# ACT vs RepeatRNN





# Experiment: One-hot addition



# Experiment: One-hot addition

Model	Task solved	Training steps	Average repetitions
LSTM	No	-	1.00
ACT-LSTM, $\tau = 10^{-1}$	No	-	1.01
ACT-LSTM, $\tau = 10^{-2}$	Yes	899 k	5.08
ACT-LSTM, $\tau = 5 \cdot 10^{-3}$	Yes	988 k	6.74
ACT-LSTM, $\tau = 10^{-3}$	No	-	11.91
Repeat-LSTM, $\rho = 2$	No	-	2.00
Repeat-LSTM, $\rho = 3$	Yes	997 k	3.00
Repeat-LSTM, $\rho = 5$	Yes	514 k	5.00
Repeat-LSTM, $\rho = 8$	Yes	576 k	8.00

# Experiment: One-hot addition

Model	Task solved	Training steps	Average repetitions
LSTM	No	-	1.00
ACT-LSTM, $\tau = 10^{-1}$	No	-	1.01
ACT-LSTM, $\tau = 10^{-2}$	Yes	899 k	5.08
ACT-LSTM, $\tau = 5 \cdot 10^{-3}$	Yes	988 k	6.74
ACT-LSTM, $\tau = 10^{-3}$	No	-	11.91
Repeat-LSTM, $\rho = 2$	No	-	2.00
Repeat-LSTM, $\rho = 3$	Yes	997 k	3.00
Repeat-LSTM, $\rho = 5$	Yes	514 k	5.00
Repeat-LSTM, $\rho = 8$	Yes	576 k	8.00

Interpretable  
hyperparameter

# Experiment: One-hot addition

Model	Task solved	Training steps	Average repetitions
LSTM	No	-	1.00
ACT-LSTM, $\tau = 10^{-1}$	No	-	1.01
ACT-LSTM, $\tau = 10^{-2}$	Yes	899 k	5.08
ACT-LSTM, $\tau = 5 \cdot 10^{-3}$	Yes	988 k	6.74
ACT-LSTM, $\tau = 10^{-3}$	No	-	11.91
Repeat-LSTM, $\rho = 2$	No	-	2.00
Repeat-LSTM, $\rho = 3$	Yes	997 k	3.00
Repeat-LSTM, $\rho = 5$	Yes	514 k	5.00
Repeat-LSTM, $\rho = 8$	Yes	576 k	8.00

Faster training....

# Experiment: One-hot addition

Model	Task solved	Training steps	Average repetitions
LSTM	No	-	1.00
ACT-LSTM, $\tau = 10^{-1}$	No	-	1.01
ACT-LSTM, $\tau = 10^{-2}$	Yes	899 k	5.08
ACT-LSTM, $\tau = 5 \cdot 10^{-3}$	Yes	988 k	6.74
ACT-LSTM, $\tau = 10^{-3}$	No	-	11.91
Repeat-LSTM, $\rho = 2$	No	-	2.00
Repeat-LSTM, $\rho = 3$	Yes	997 k	3.00
Repeat-LSTM, $\rho = 5$	Yes	514 k	5.00
Repeat-LSTM, $\rho = 8$	Yes	576 k	8.00

...and faster inference



# Open Science

[AUTHORS](#)[PAPER](#)[ACT MODEL](#)[REPEAT-RNN](#)[RESULTS](#)[SLIDES](#)[CODE](#)[ACKNOWLEDGEMENTS](#)

Fork me on GitHub

## Comparing Fixed and Adaptive Computation Time for Recurrent Neural Networks



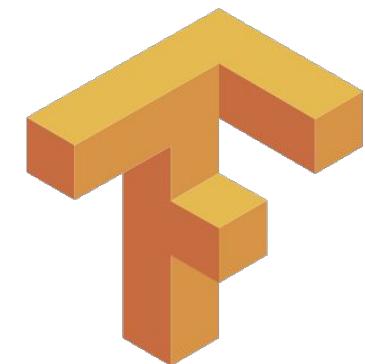
Daniel Fojo



Xavier Giró-i-Nieto



Víctor Campos

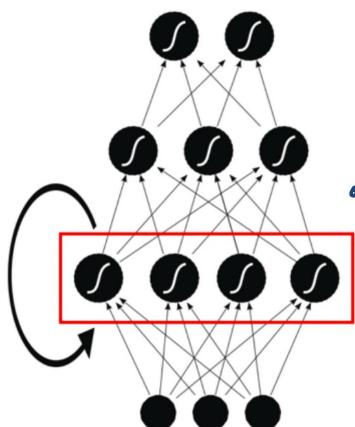


[bit.ly/repeat-rnn](https://bit.ly/repeat-rnn)

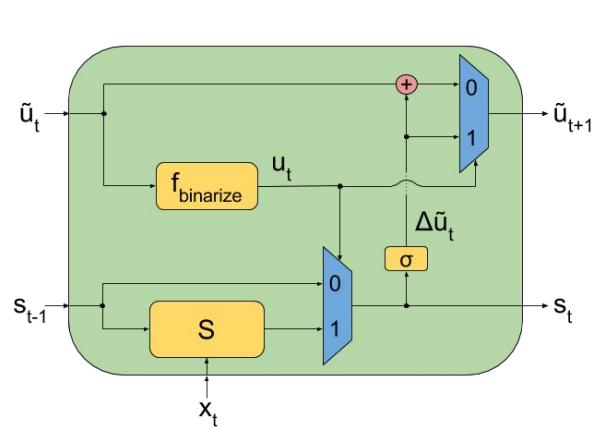


# Outline

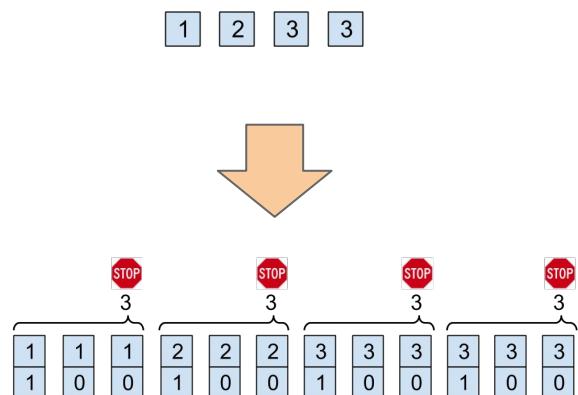
Recurrent Neural Networks (RNNs)



SkipRNN  
[ICLR 2018]



RepeatRNN  
[ICLRW 2018]



# Acknowledgements



**Barcelona  
Supercomputing  
Center**

*Centro Nacional de Supercomputación*



**EXCELENCIA  
SEVERO  
OCHOA**



**Generalitat de Catalunya  
Departament d'Economia  
i Coneixement**



GOBIERNO  
DE ESPAÑA

MINISTERIO  
DE ECONOMÍA  
Y COMPETITIVIDAD

# Deep Learning courses @ UPC (videos)

## DEEP LEARNING FOR ARTIFICIAL INTELLIGENCE

videos will be online

Master Course UPC ETSETB TelecomBCN Barcelona. Autumn 2017.



### Instructors



### Organizers



Supporters



aws Educate

GitHub Education

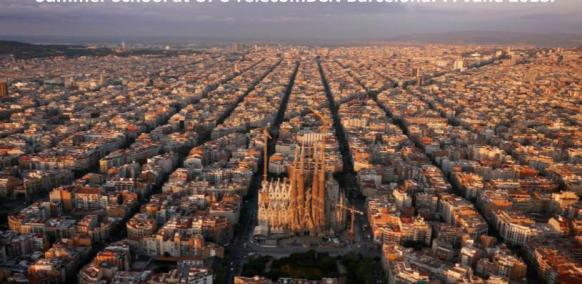
+ info: <http://dlai.deeplearning.barcelona>

- [MSC course](#) (2017)
- [BSc course](#) (2018)

Next edition Autumn 2018

## DEEP LEARNING FOR COMPUTER VISION

Summer School at UPC TelecomBCN Barcelona. ?? June 2018.



### Instructors



### Organized by



UNIVERSITAT POLITÈCNICA  
DE CATALUNYA  
BARCELONATECH



Supported by



GitHub Education

+ info: <http://bit.ly/dlcv2018>

- [1st edition](#) (2016)
- [2nd edition](#) (2017)
- [3rd edition](#) (2018)

Summer School (starts 28/06)

## DEEP LEARNING FOR SPEECH AND LANGUAGE

Winter School at UPC TelecomBCN Barcelona. 24-30 January 2018.



### Instructors



### Organized by



UNIVERSITAT POLITÈCNICA  
DE CATALUNYA  
BARCELONATECH



Supported by



GitHub Education

+ info: <https://telecombcn-dl.github.io/2018-dsl/>

- [1st edition](#) (2017)
- [2nd edition](#) (2018)

Next edition Winter/Spring 2019



# Xavier Giro-i-Nieto



UNIVERSITAT POLITÈCNICA  
DE CATALUNYA  
BARCELONATECH

Comment slides from:  
<http://bit.ly/InsightDL2018>



The slide is for the REWORK Deep Learning in Retail & Advertising Summit, held in London, UK on 16th March 2018. It features a yellow header with the REWORK logo, a blue footer with the UPC logo, and a central white area with the title "Learning Where and When to Look". It includes links to "Slides on GDrive" and "@DocXavi" on Twitter, and a "#REWORKDL" hash tag. Below the title is a photo of Xavier Giro-i-Nieto and his contact information: Xavier Giro-i-Nieto, Associate Professor at Universitat Politècnica de Catalunya, Technical University of Catalonia.



#InsightDL2018

We are looking for both  
industrial & academic partners:



xavier.giro@upc.edu



@DocXavi

# Questions?