Insight
Centre for Data Analytics

DEEP LEARNING WORKSHOP

Dublin City University
21-22 May 2018

DCU

#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 Politecnica de Catalunya (UPC)

bit.ly/InsightDL2018
#InsightDL2018

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

DCU Dublin City University Ollscoil Chathair Bhaile Átha Cliath

Insight Centre for Data Analytics

Kevin McGuinness

Noel E. O'Connor

Cathal Gurrin

Eva Mohedano

COLUMBIA UNIVERSITY

Shih-Fu Chang

Francisco Roldan

Marta Coll

Alejandro Woodward

Paula Gómez

Marc Górriz

UOC Universitat Oberta de Catalunya

Carles Ventura

Our team @ industry

VILYNX

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH
Robotics Innovation Center

MEDIAPRO

crisalix
VIRTUAL AESTHETICS

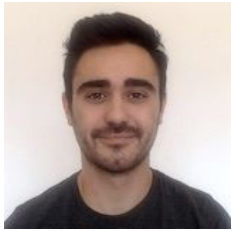
Elisenda Bou

Dèlia Fernández

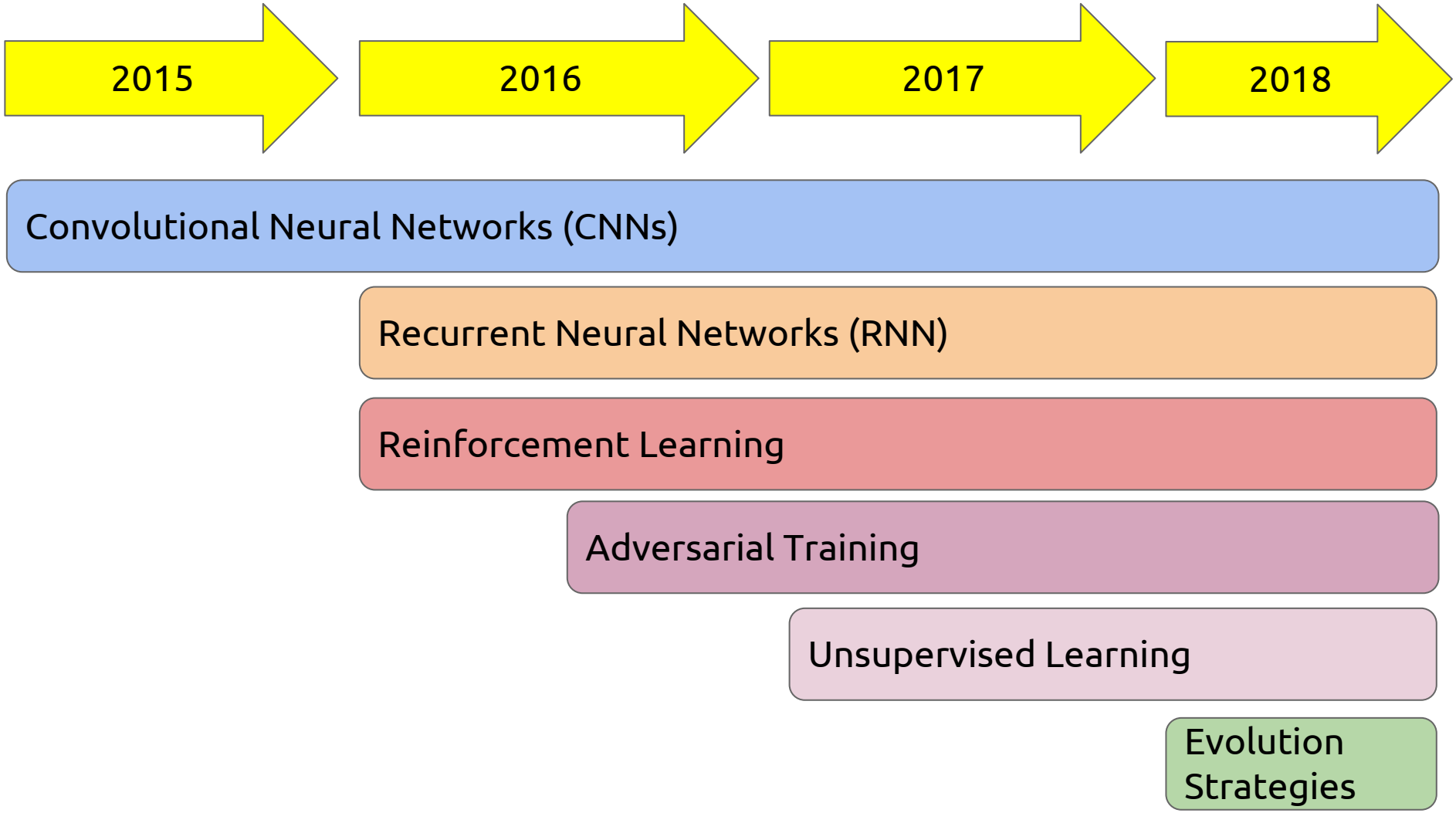
Sebastian Palacio

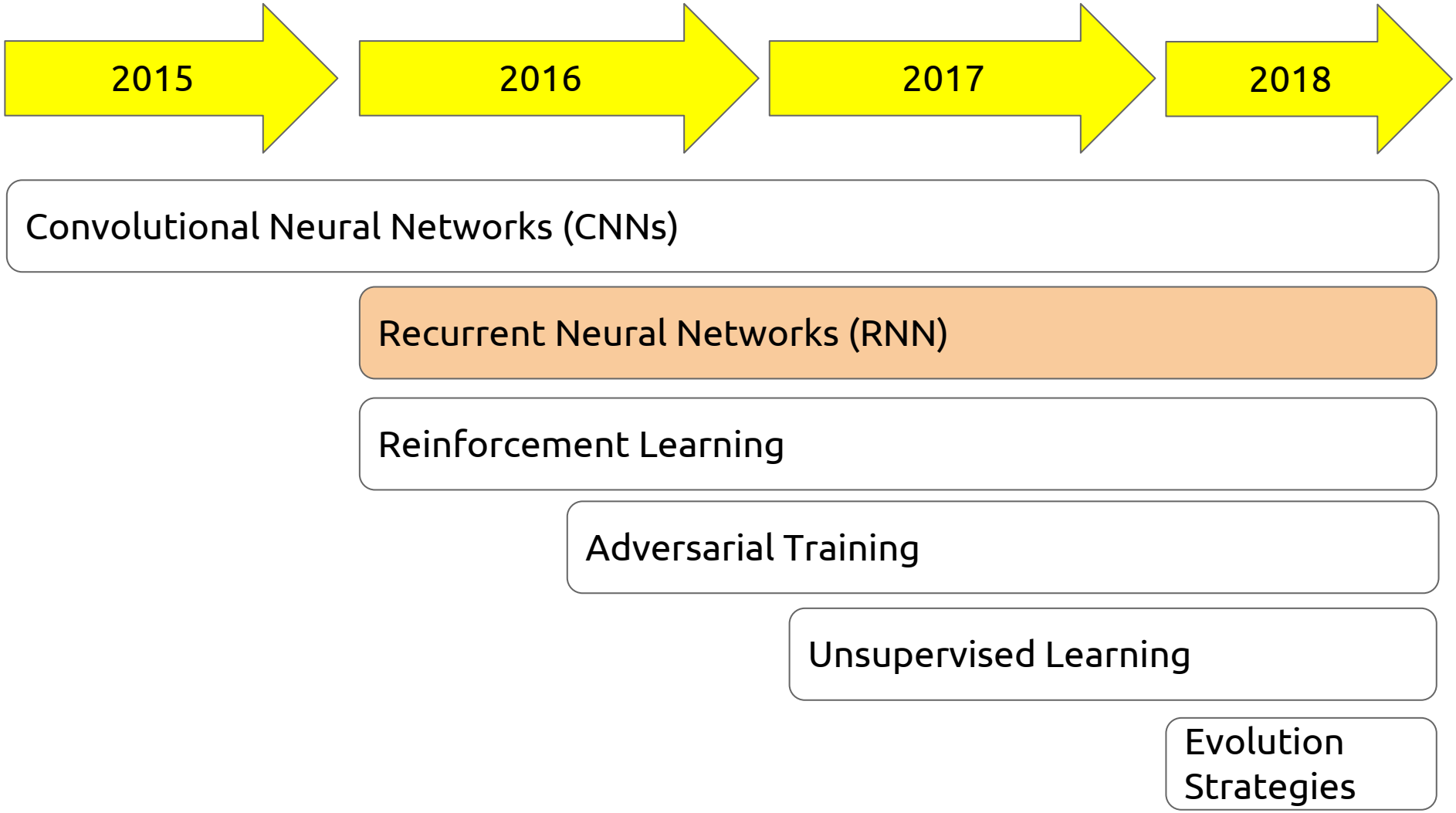
Carlos Arenas

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: Dynamic Computation

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

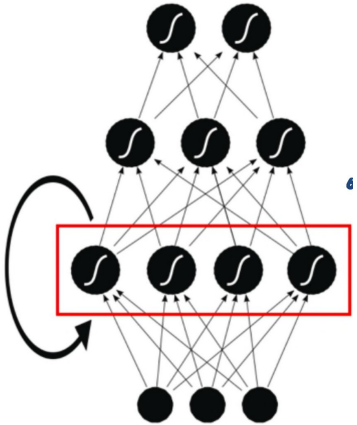
2+2

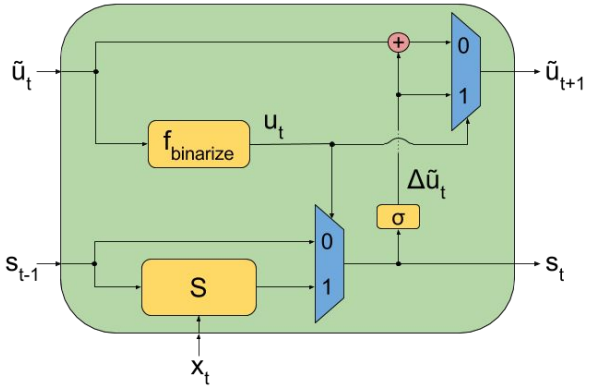
Outline

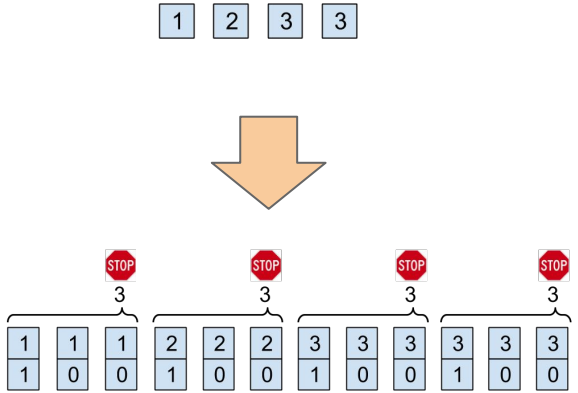
Recurrent Neural Networks (RNNs)

SkipRNN [ICLR 2018]

RepeatRNN [ICLRW 2018]





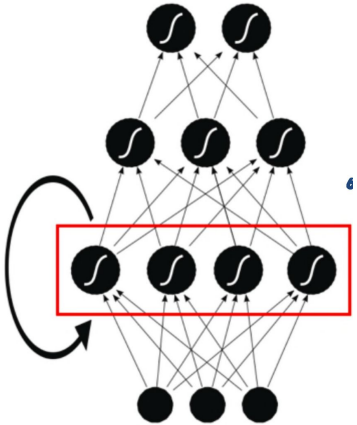


Outline

Recurrent Neural Networks (RNNs)

SkipRNN [ICLR 2018]

RepeatRNN [ICLRW 2018]



Feed-forward Neural Network

Feed Forward

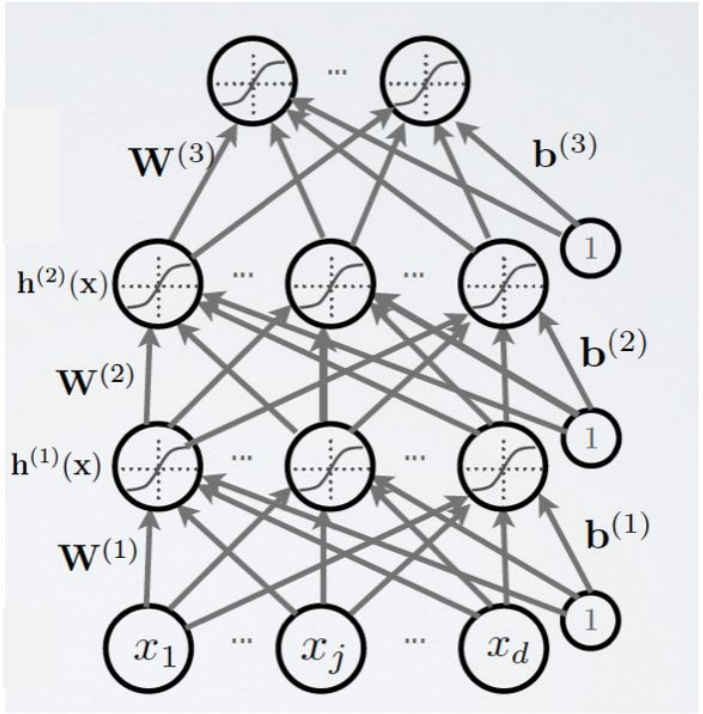


Figure: Hugo Larochelle

OUTPUT (y)

Feed-forward Weights (W_i)

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

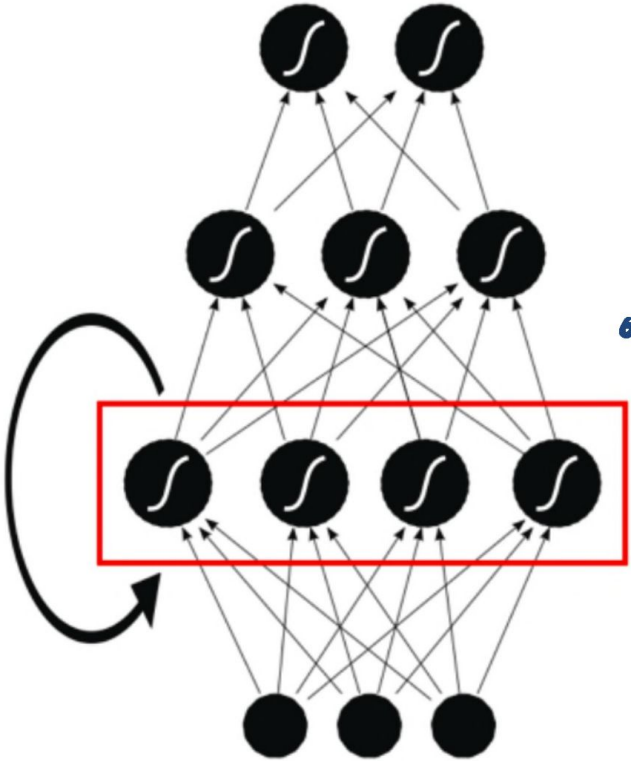
Hidden States h_1 & h_2

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

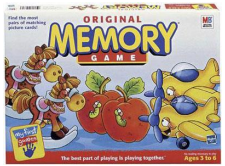
$$h_1 = f(W_1 \cdot x + b_1)$$

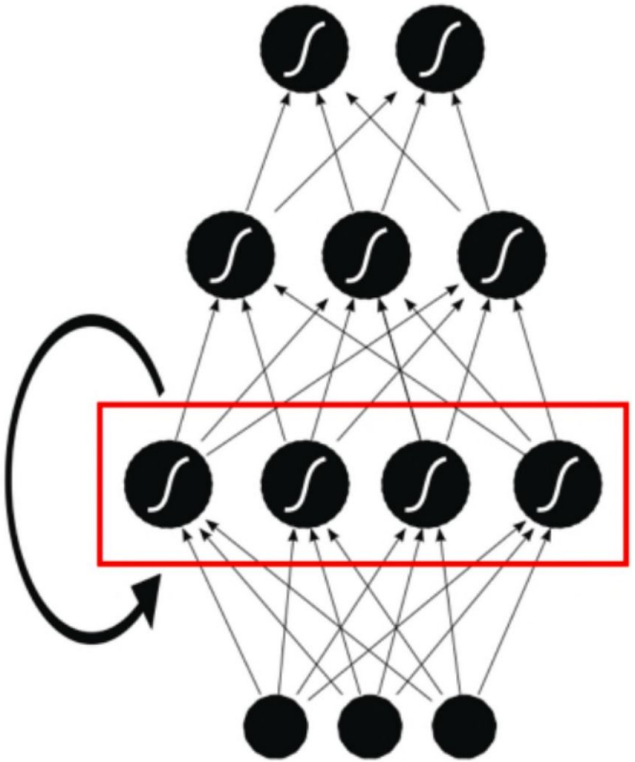
INPUT (x)

Recurrent Neural Network



Recurrent Neural Network



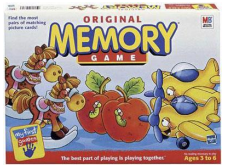


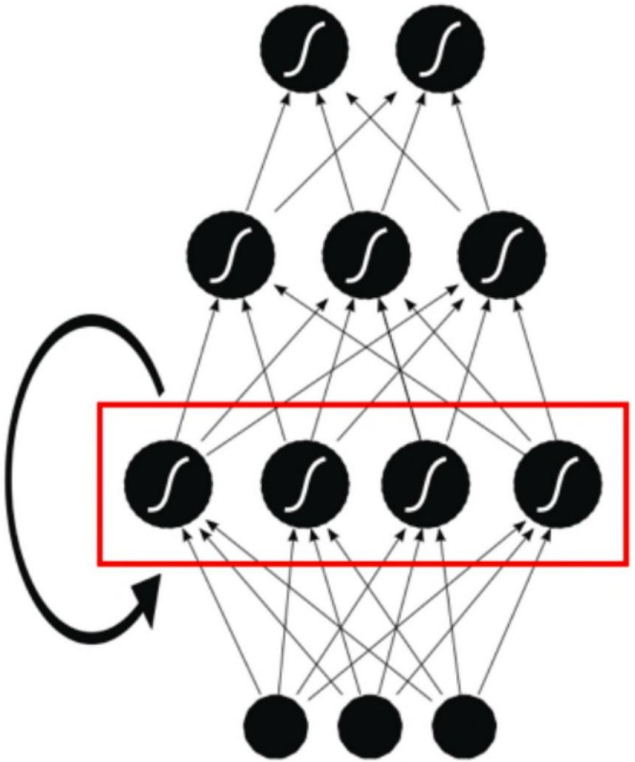
Feed-forward Weights (W)

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

Recurrent Weights (U)

Recurrent Neural Network





Updated state

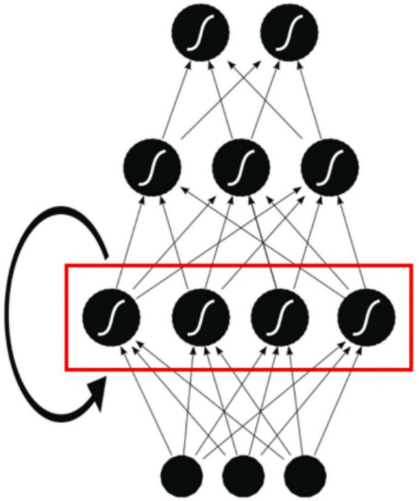
Previous state

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

INPUT (x)

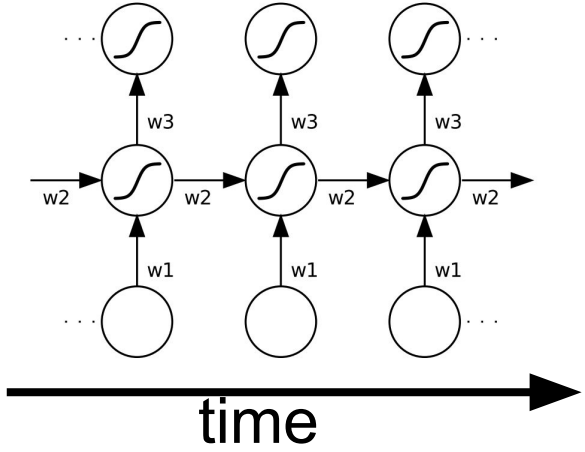
Recurrent Neural Network

Unfold
(Rotation 90º)

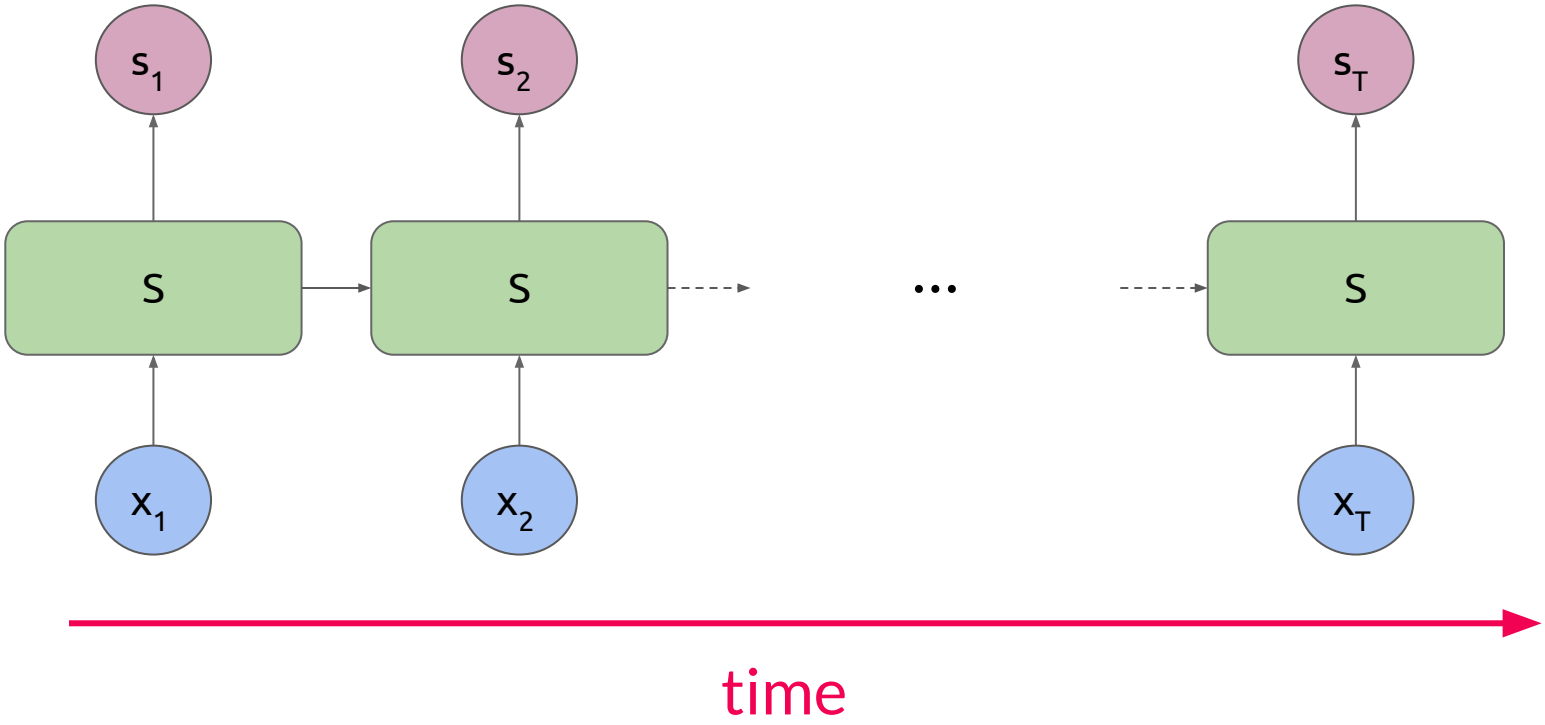


time

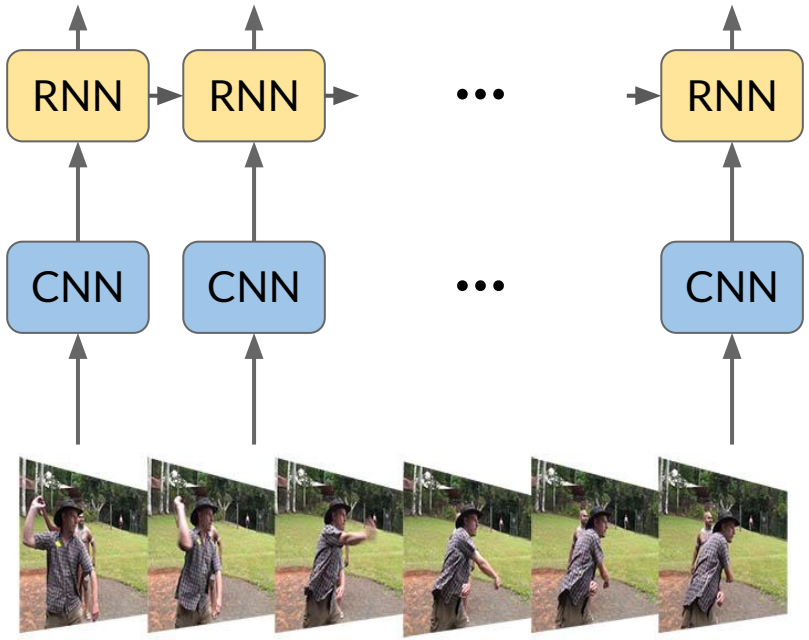
Unfold
(Rotation 90º)



Recurrent Neural Network

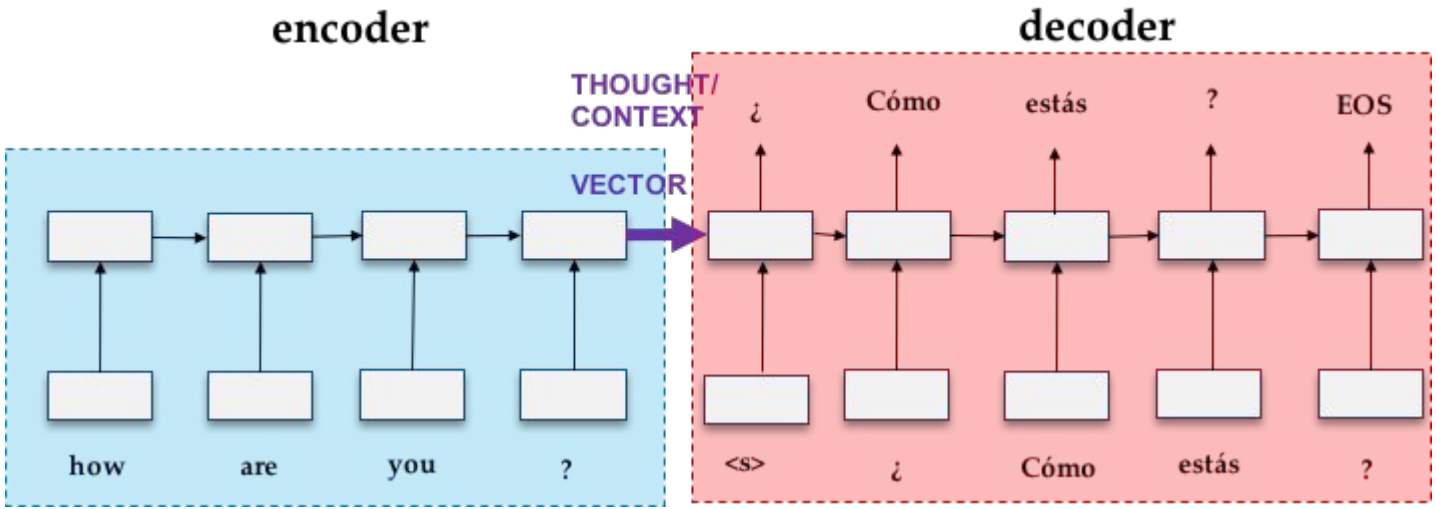


Recurrent Neural Network



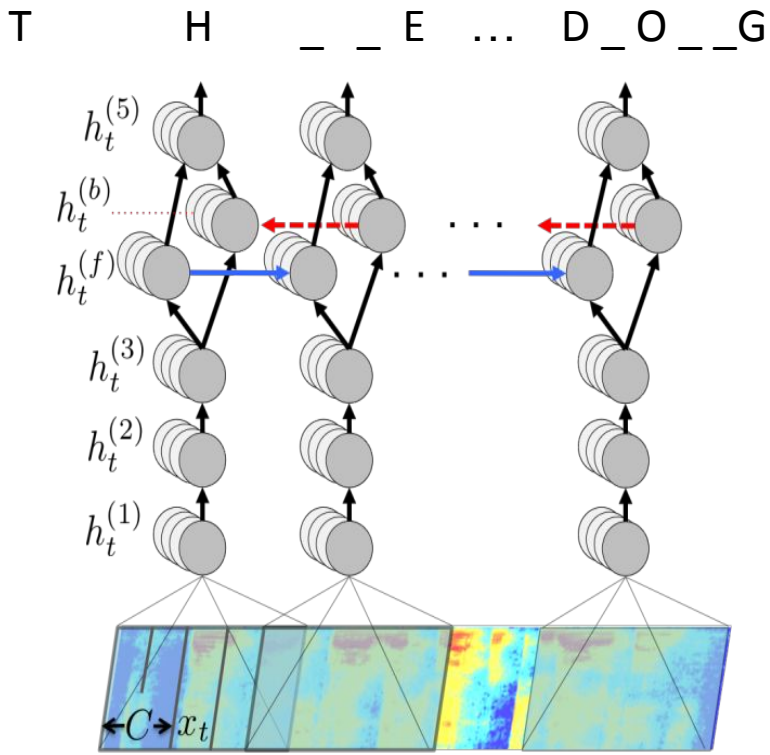
Video sequences (eg. action recognition)

Recurrent Neural Network



Word sequences (eg. Machine Translation)

Recurrent Neural Network



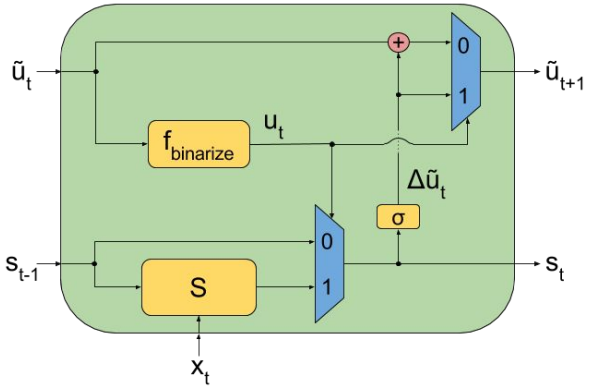
Spectrograms sequence (eg. speech recognition)

Outline

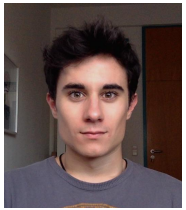
Recurrent Neural Networks (RNNs)

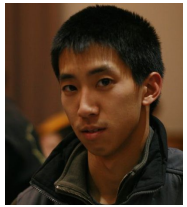
SkipRNN [ICLR 2018]

RepeatRNN [ICLRW 2018]



Skip RNN: Learning to Skip State Updates in RNNs


Víctor Campos


Brendan Jou

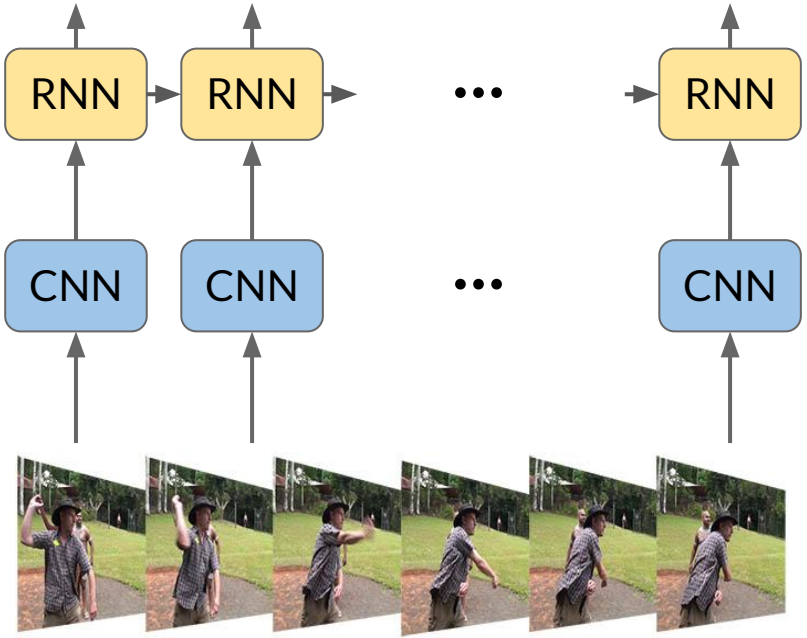

Jordi Torres


Xavier Giró-i-Nieto


Shih-Fu Chang

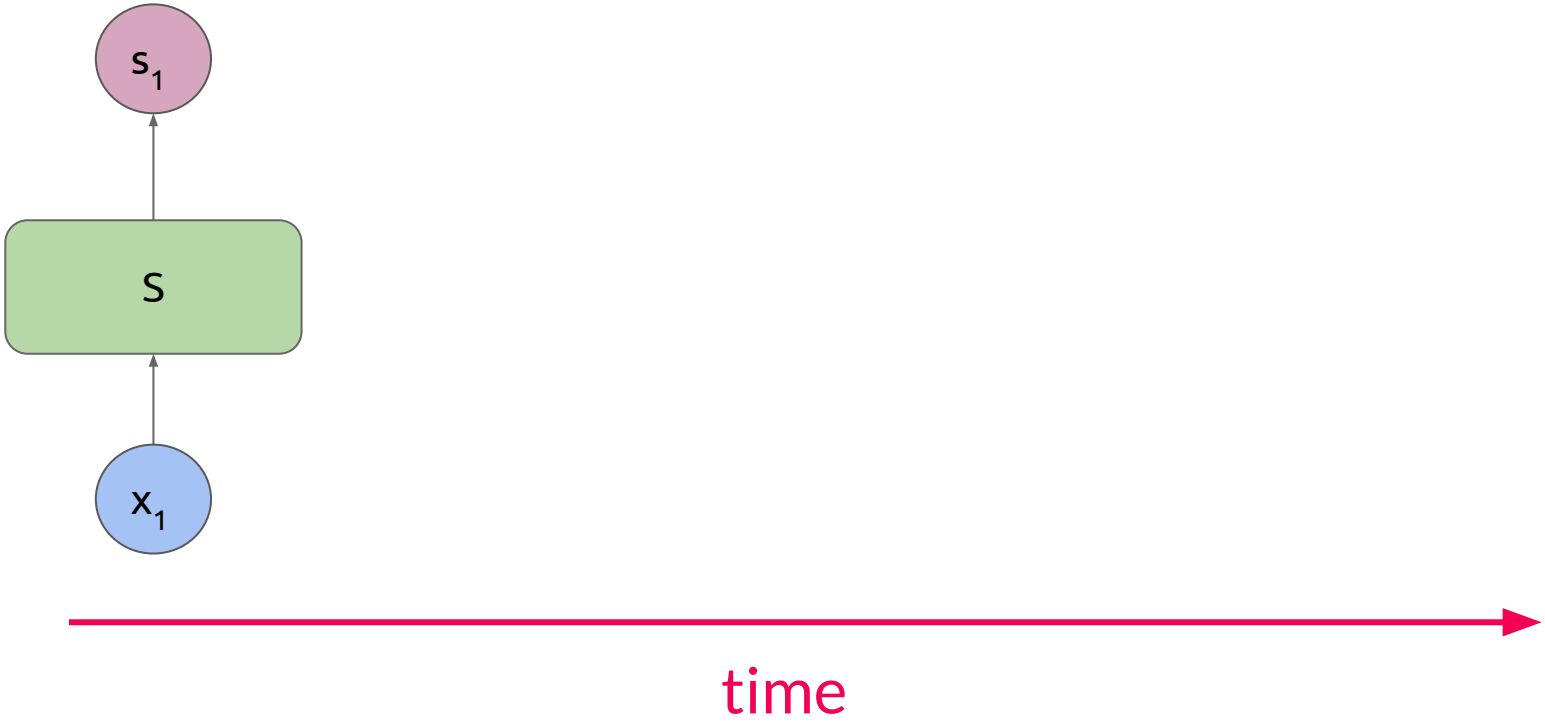
Victor Campos, Brendan Jou, Xavier Giro-i-Nieto, Jordi Torres, and Shih-Fu Chang. “Skip RNN: Learning to Skip State Updates in Recurrent Neural Networks”, ICLR 2018.

Motivation

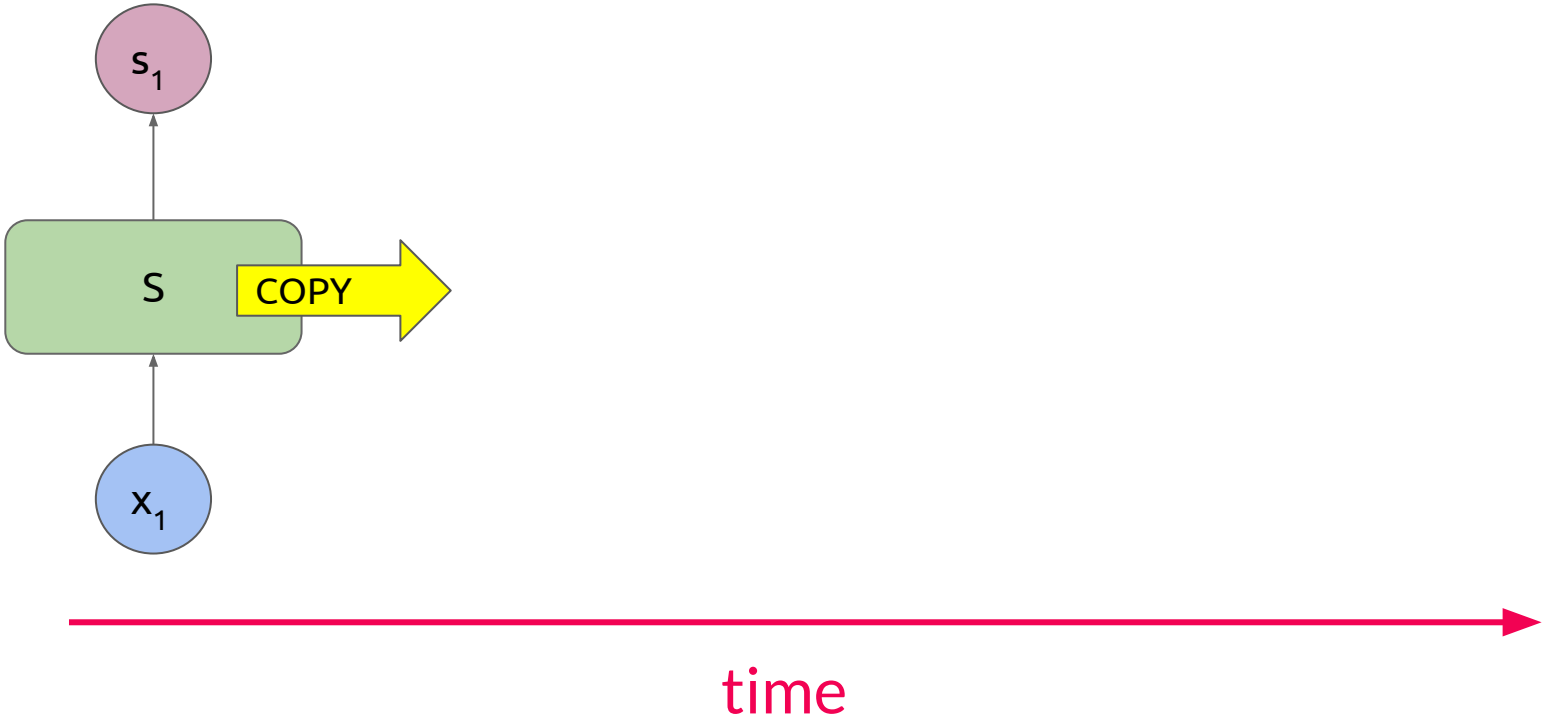


Used Unused

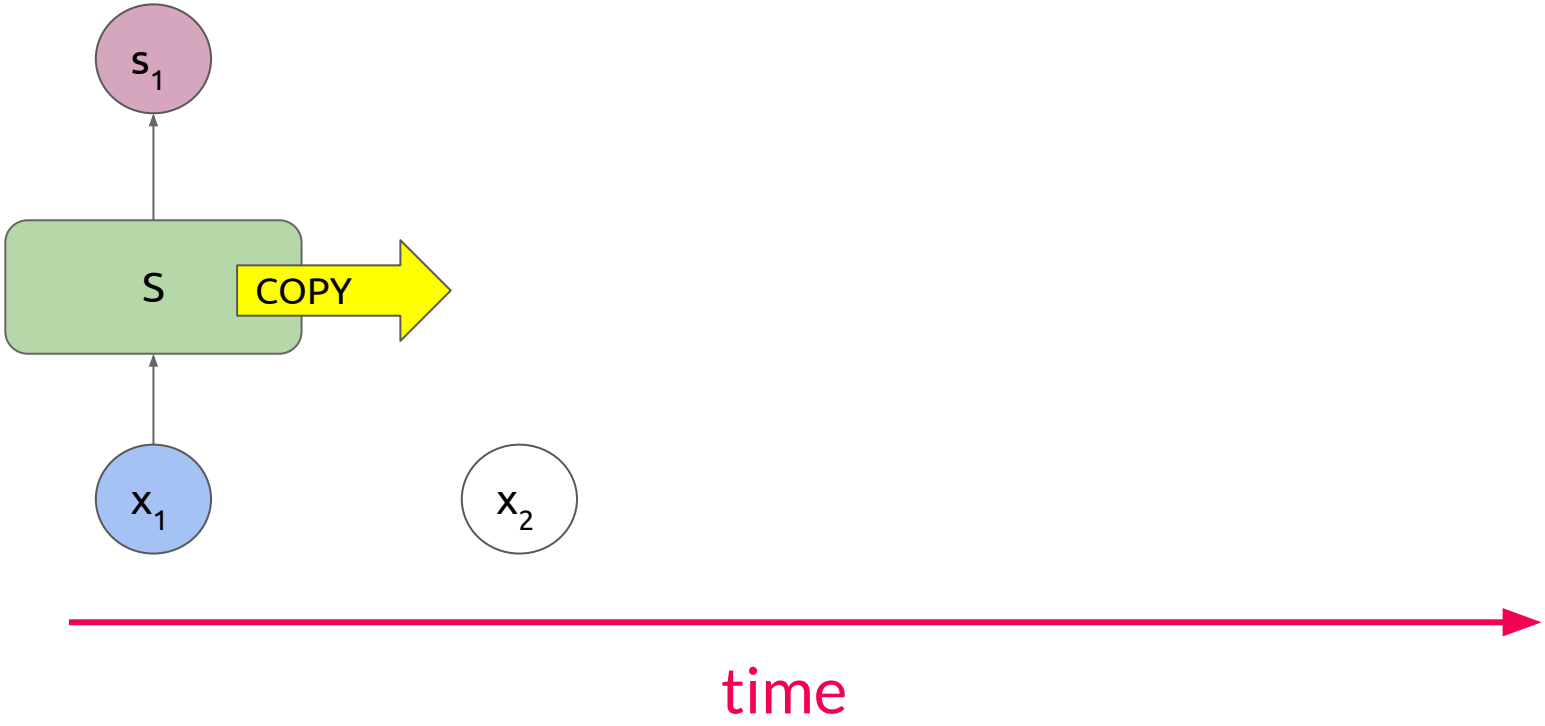
SkipRNN



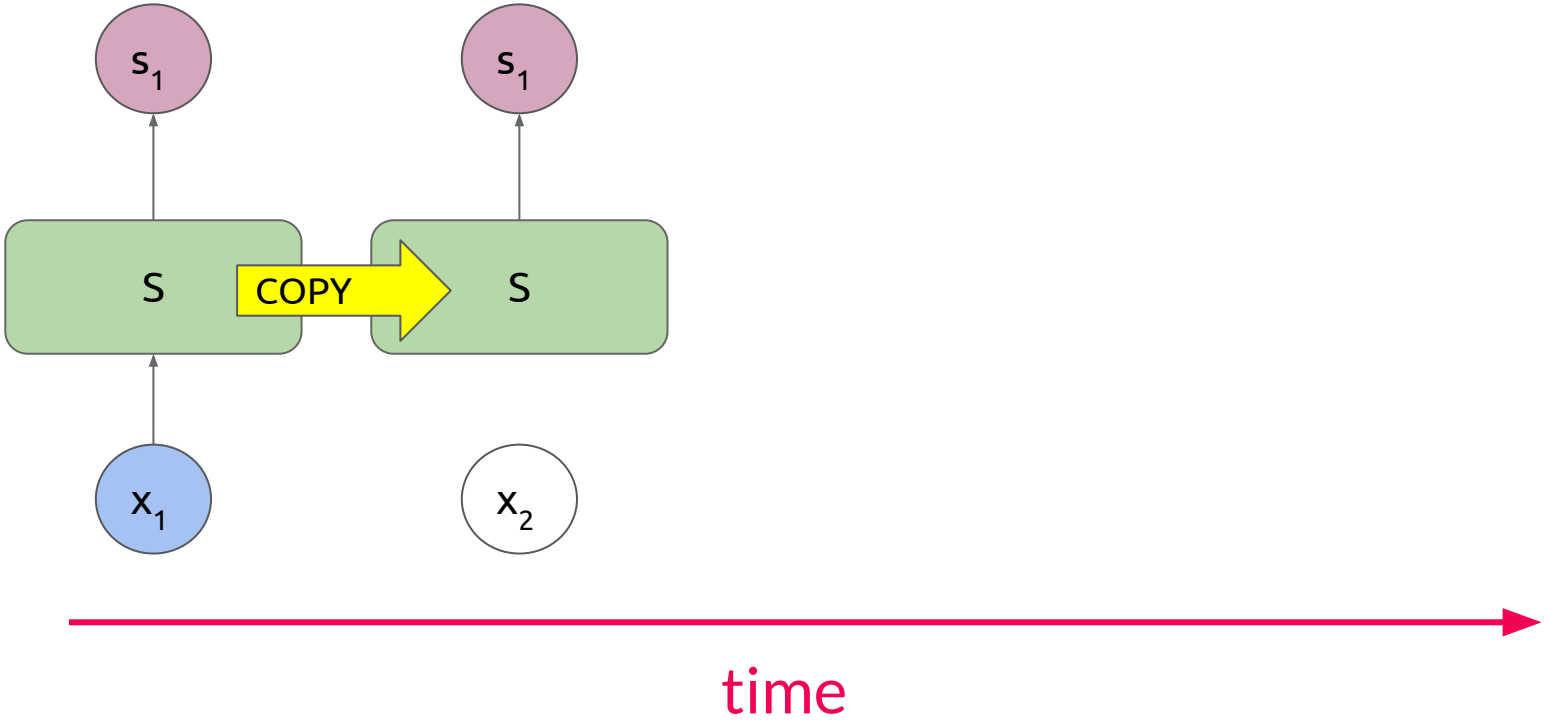
SkipRNN



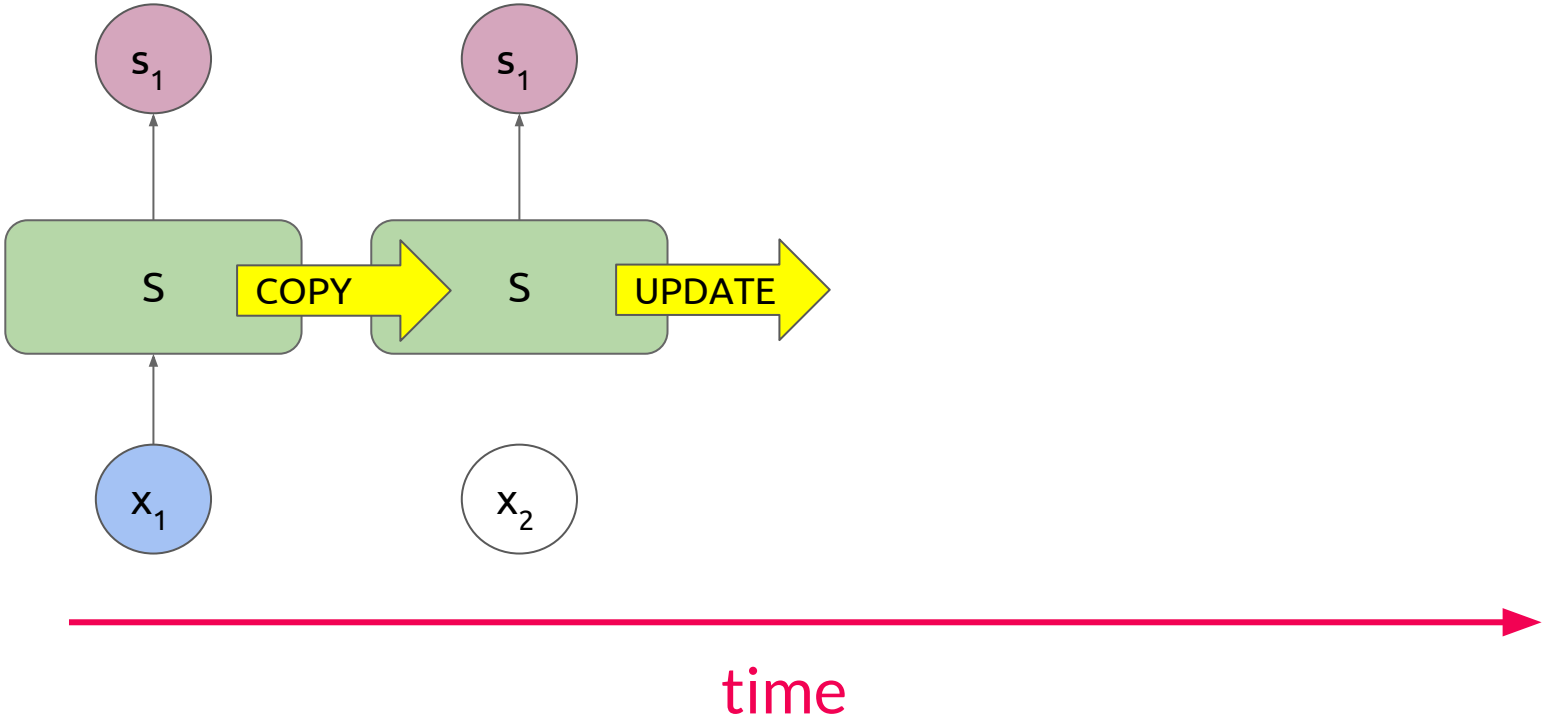
SkipRNN



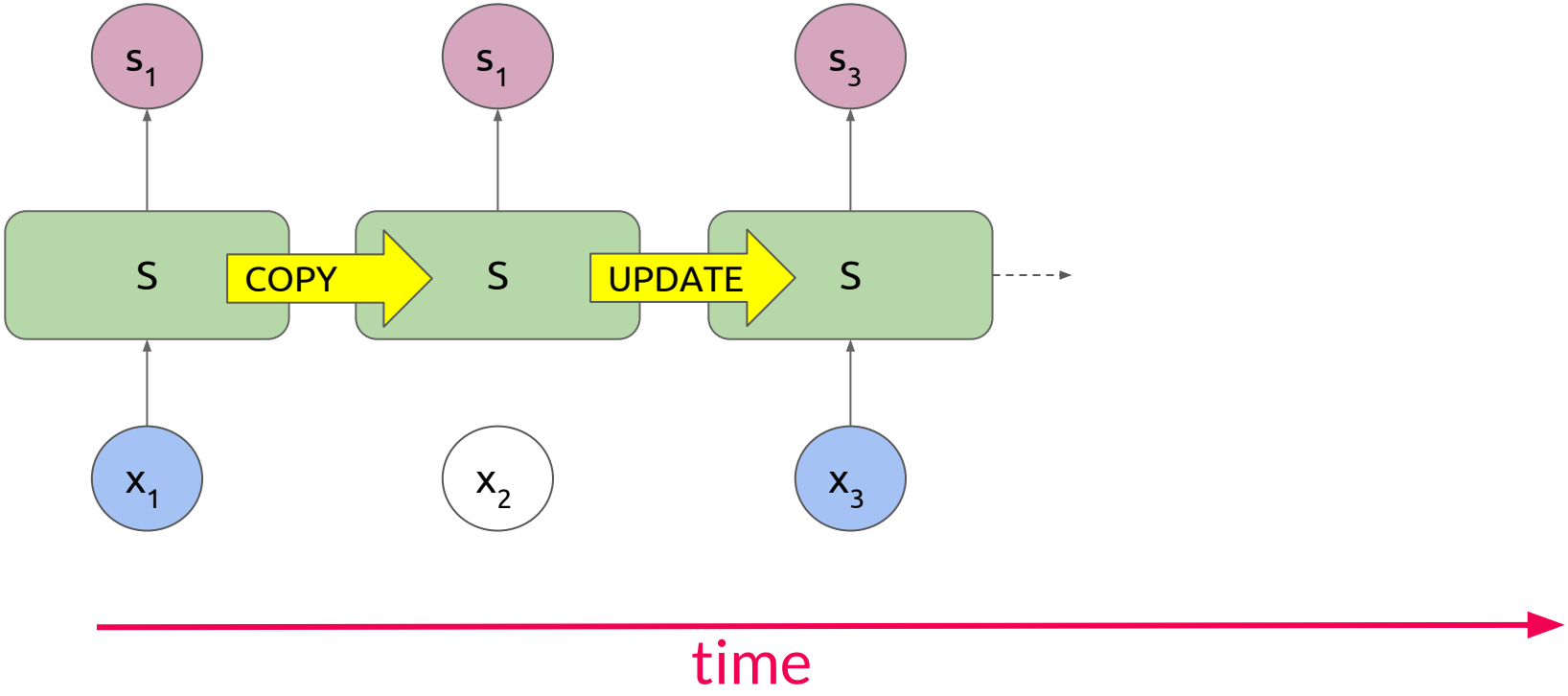
SkipRNN



SkipRNN

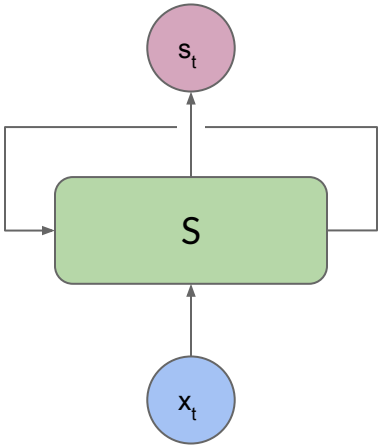


SkipRNN



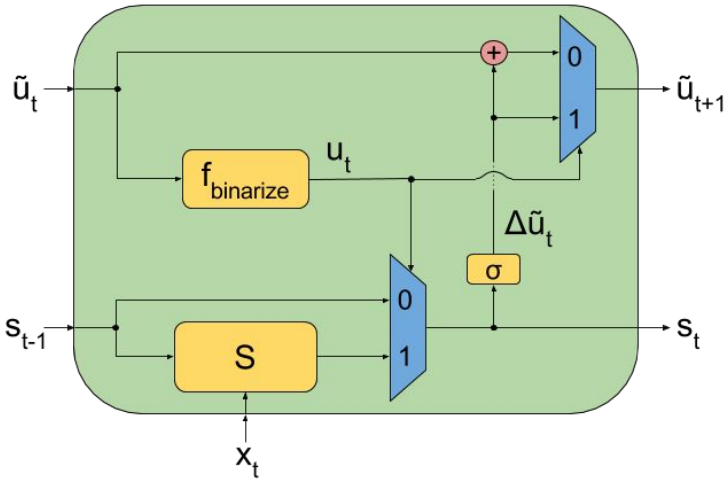
SkipRNN

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



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

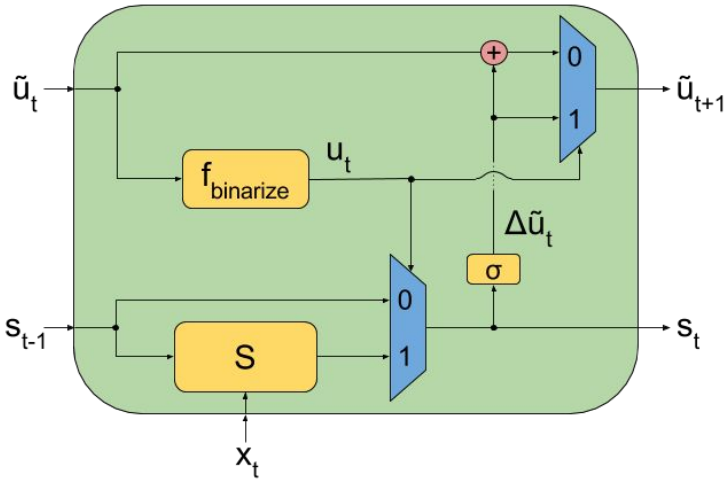
SkipRNN



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

Update state gate ∈ {0, 1}

SkipRNN

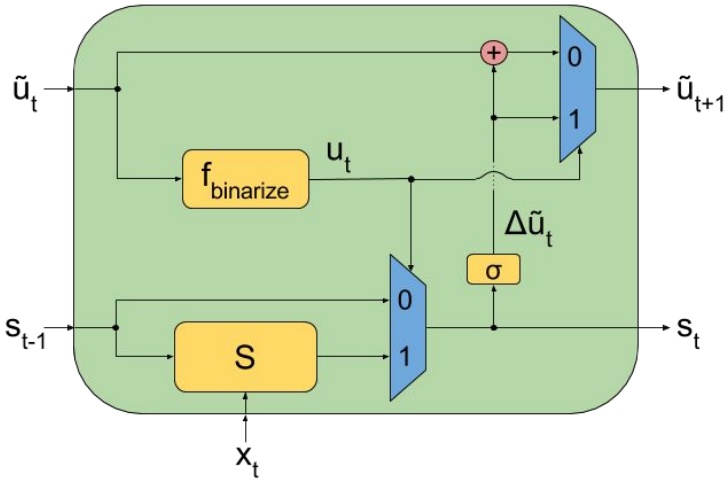


$$u_t = f_{binarize}(\tilde{u}_t)$$

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

Update state gate ∈ {0, 1}
Update state probability ∈ [0, 1]

SkipRNN



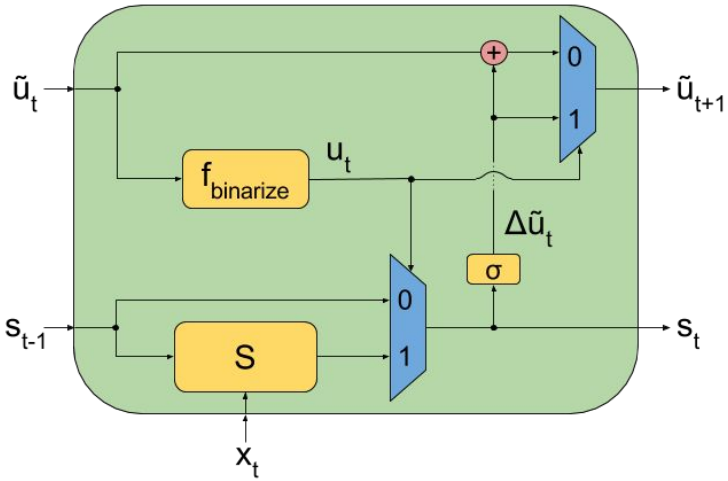
$$u_t = f_{binarize}(\tilde{u}_t)$$

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

Update state gate ∈ {0, 1}

Update state probability ∈ [0, 1]

SkipRNN



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

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

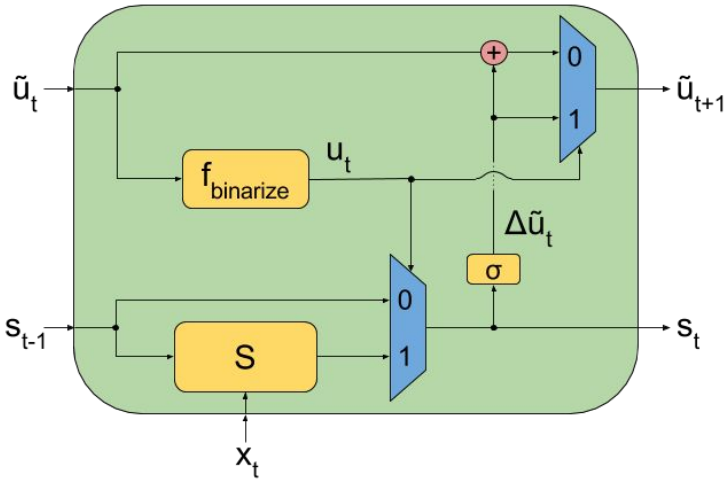
$$\tilde{u}_{t+1} = u_t \cdot \Delta\tilde{u}_t + (1 - u_t) \cdot (\tilde{u}_t + \min(\Delta\tilde{u}_t, 1 - \tilde{u}_t))$$

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

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

Increment for the update state probability

SkipRNN



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

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

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

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

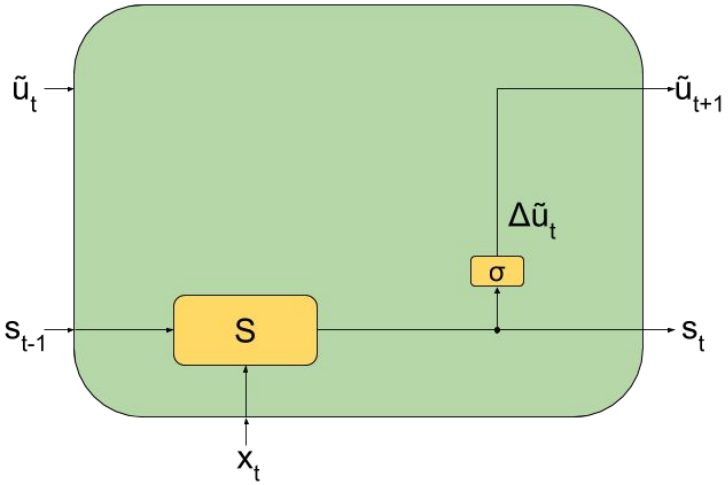
Update state gate ∈ {0, 1}

Update state probability ∈ [0, 1]

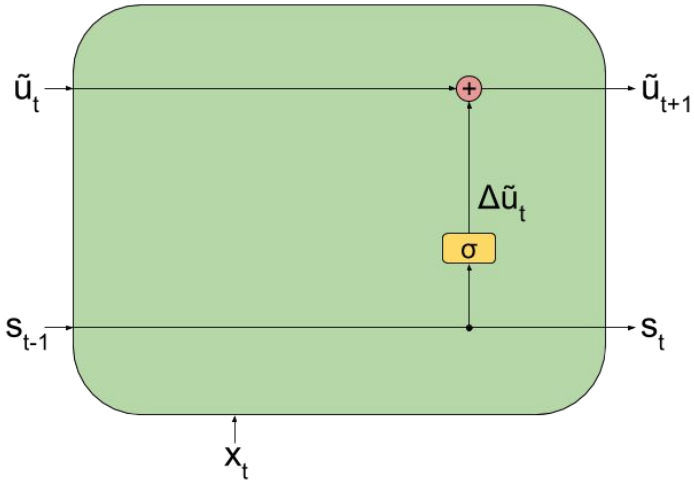
Increment for the update state probability

SkipRNN

Update state ($u_t = 1$)

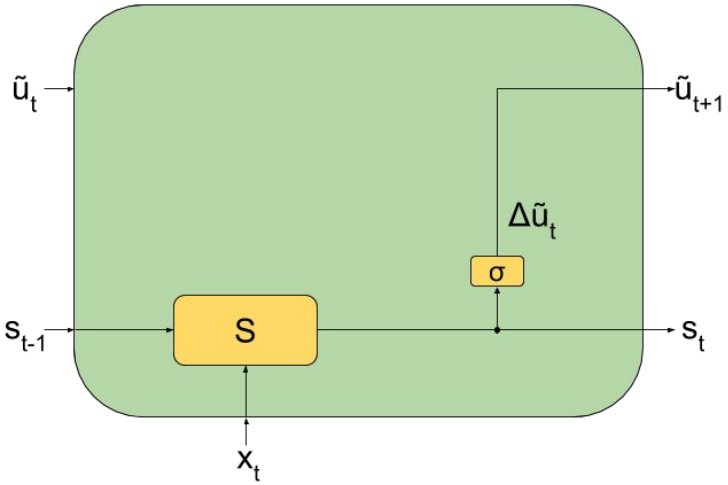


Copy state ($u_t = 0$)

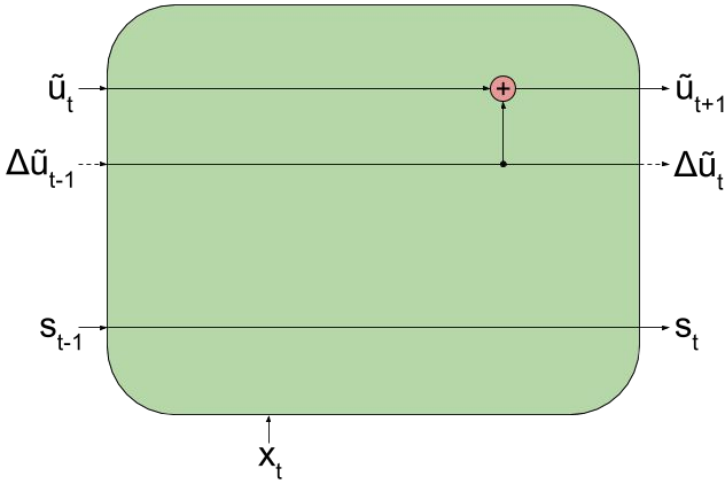


SkipRNN

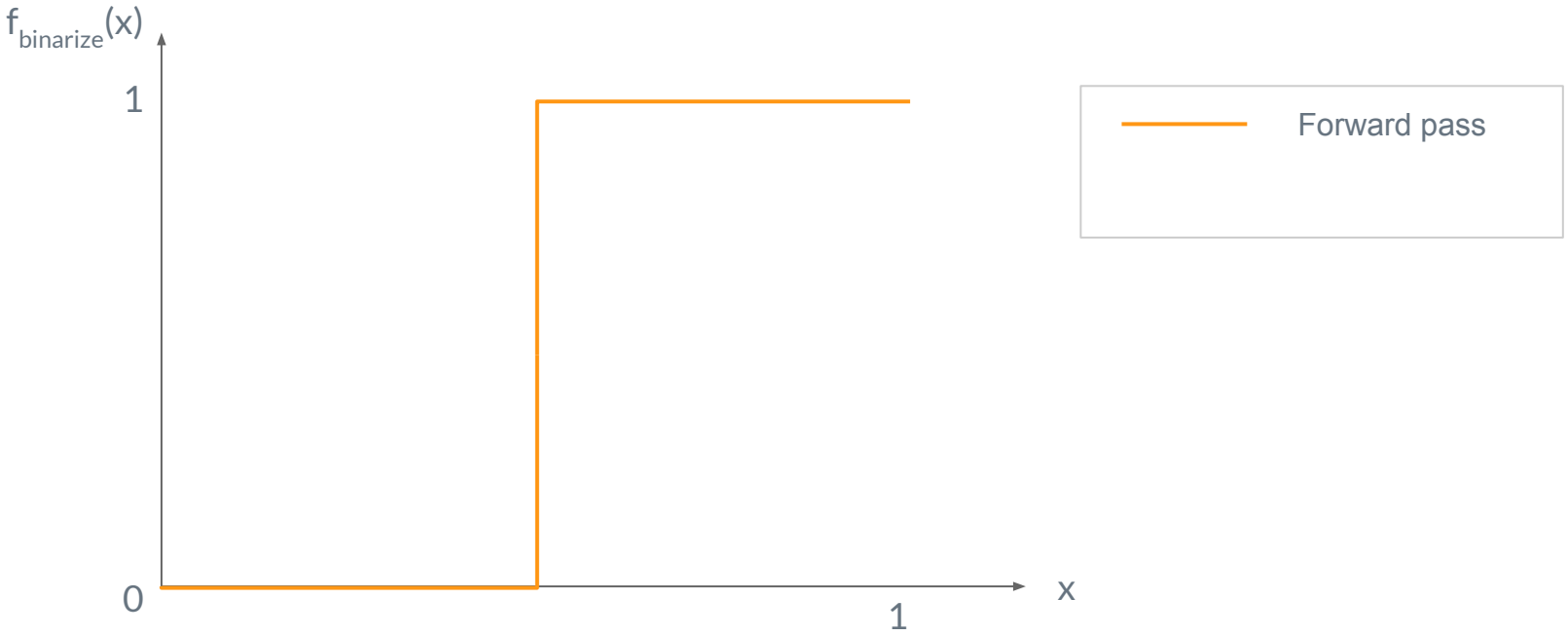
Update state ($u_t = 1$)



Copy state ($u_t = 0$)

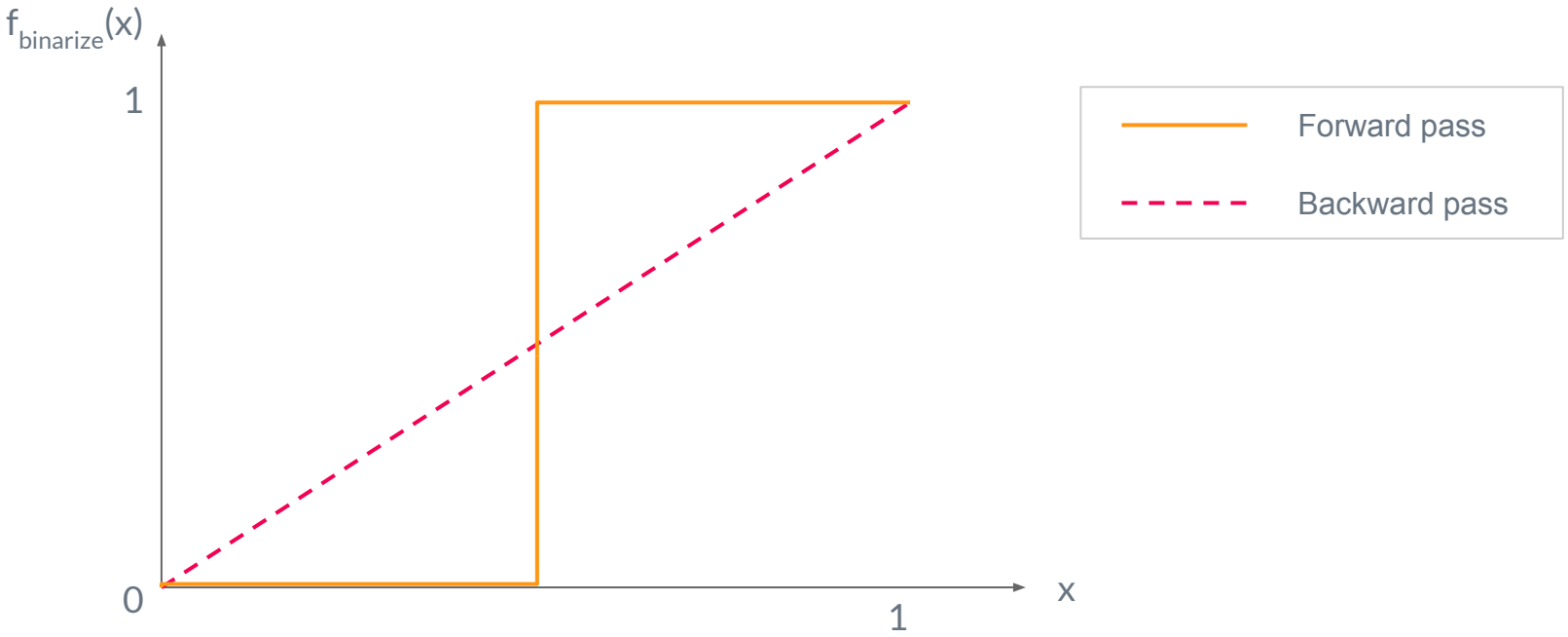


Straight Though Estimator



G. Hinton, Neural Networks for Machine Learning, Coursera video lectures 2012

Straight Though Estimator



G. Hinton, Neural Networks for Machine Learning, Coursera video lectures 2012

Limiting computation

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

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

cost per sample (hyperparameter)

Experiments: Sequential MNIST

Used
Unused

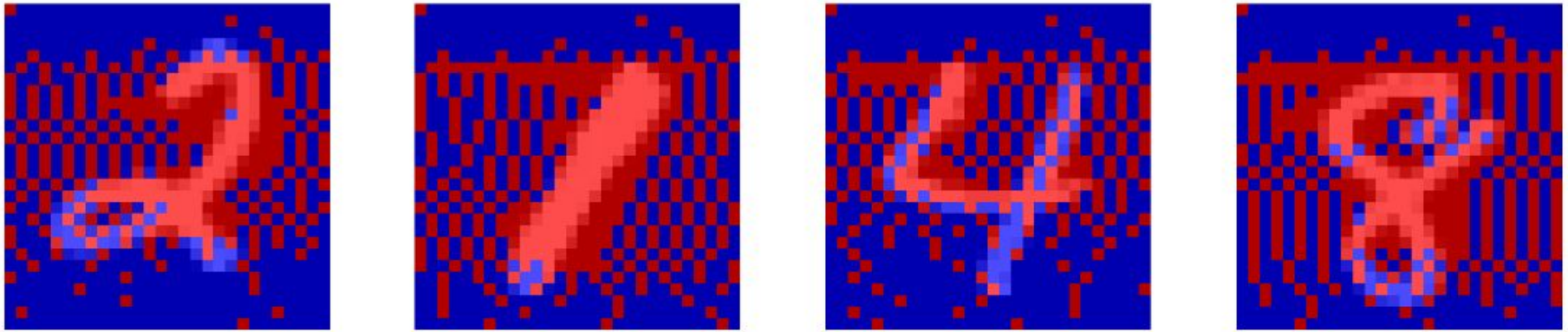
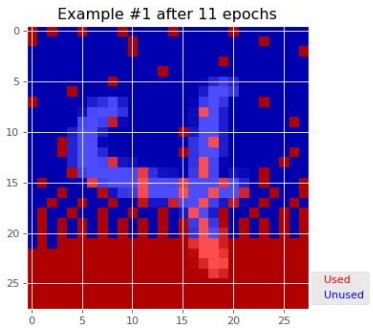


Figure 3: Sample usage examples for the Skip LSTM with $\lambda = 10^{-4}$ on the test set of MNIST. Red pixels are used, whereas blue ones are skipped.

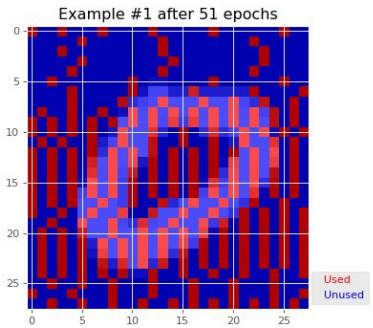
Experiments: Sequential MNIST

11 epochs



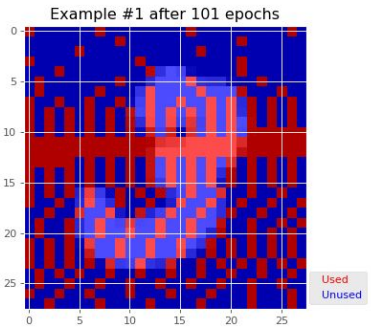
~30% acc

51 epochs



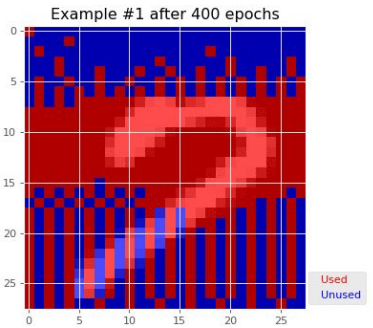
~50% acc

101 epochs



~70% acc

400 epochs



~95% acc

Epochs for Skip LSTM ($\lambda = 10^{-4}$)

Used
Unused

Experiments: Sequential MNIST

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

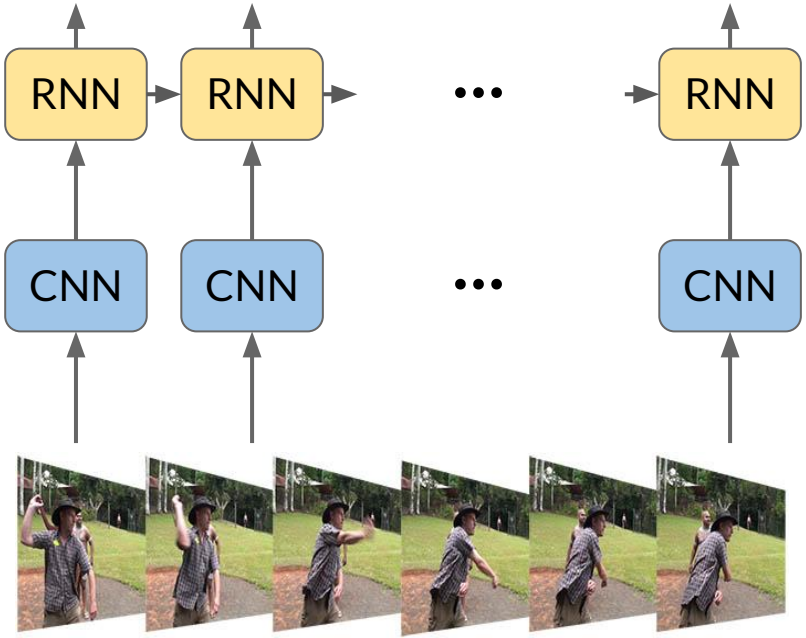
Better accuracy...

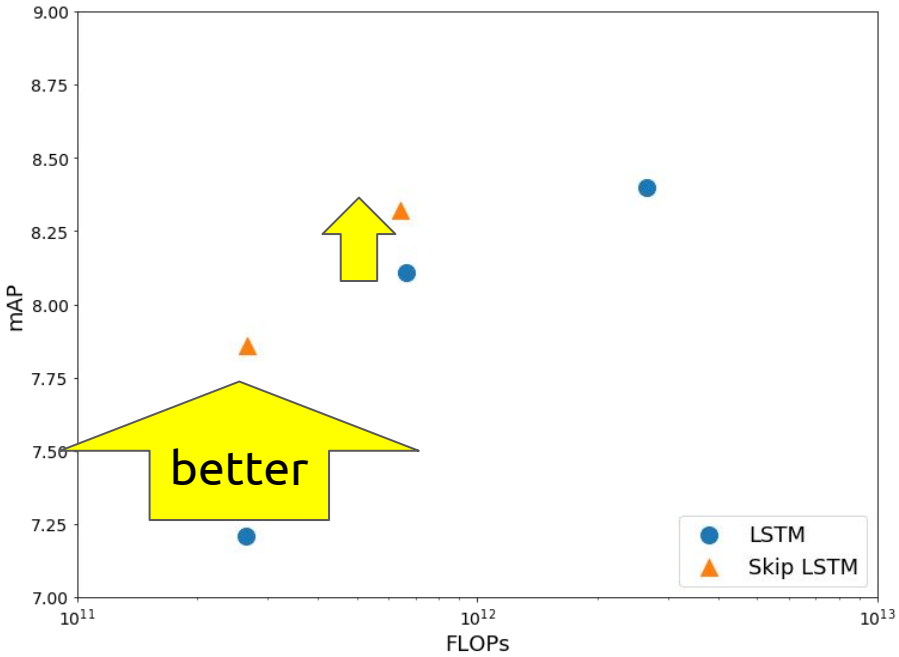
Experiments: Sequential MNIST

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

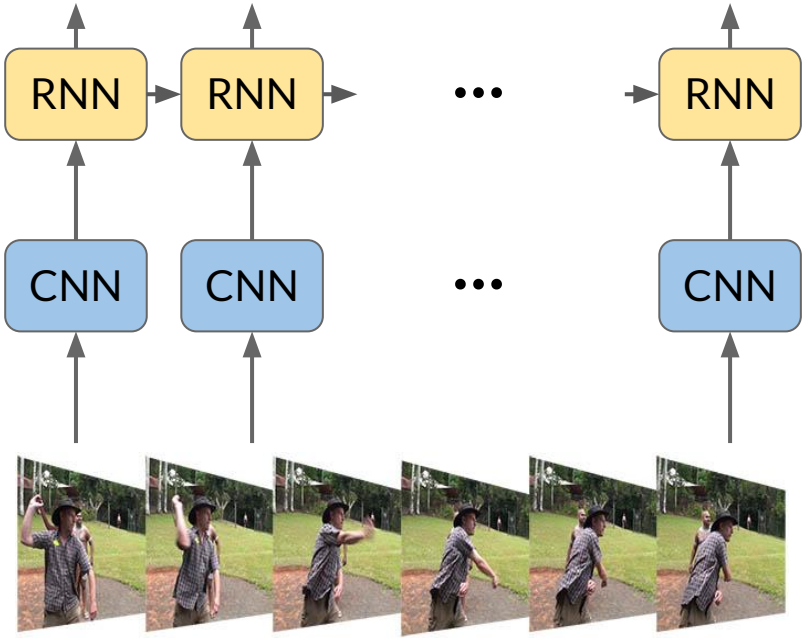
...with less computation

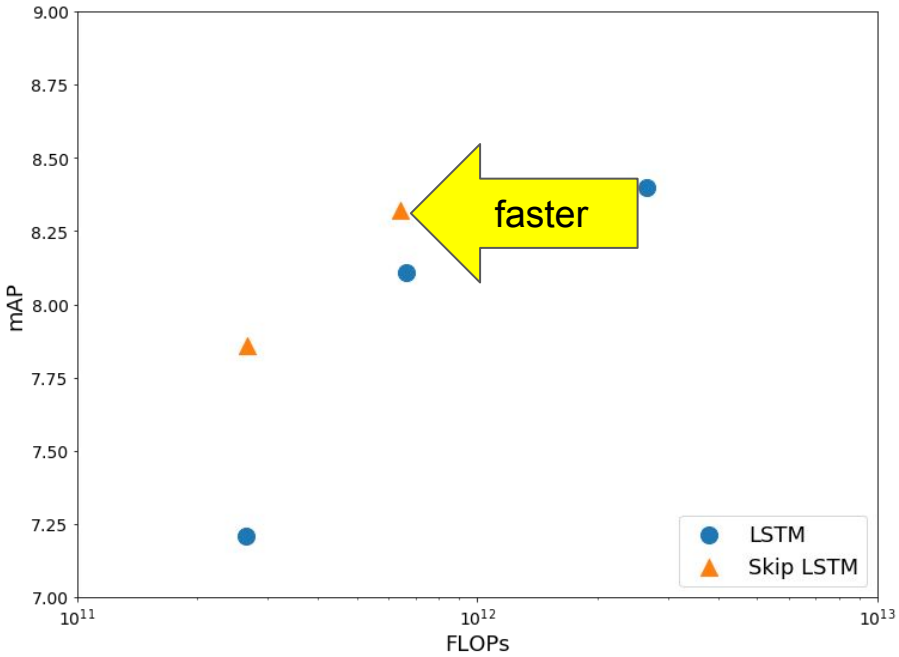
Experiments: Action Localization



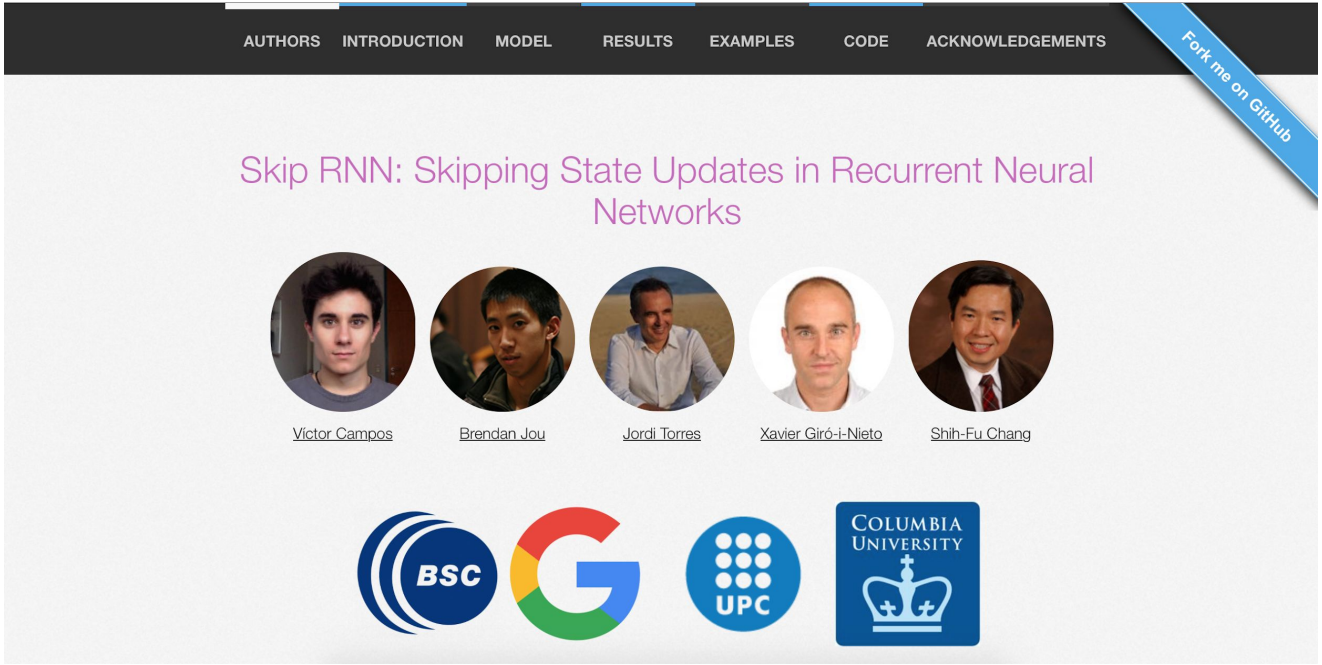


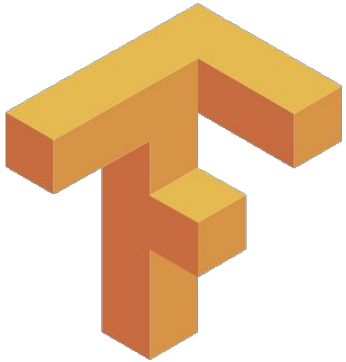
Experiments: Action Localization





Open science





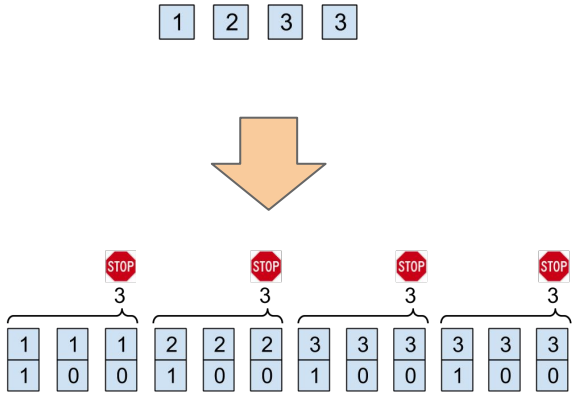
https://git.io/skip-rnn

Outline

Recurrent Neural Networks (RNNs)

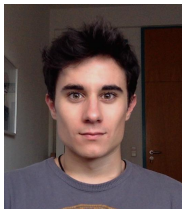
SkipRNN [ICLR 2018]

RepeatRNN [ICLRW 2018]



Reproducing and Analyzing Adaptive Computation Time in PyTorch and TensorFlow

Dani Fojo

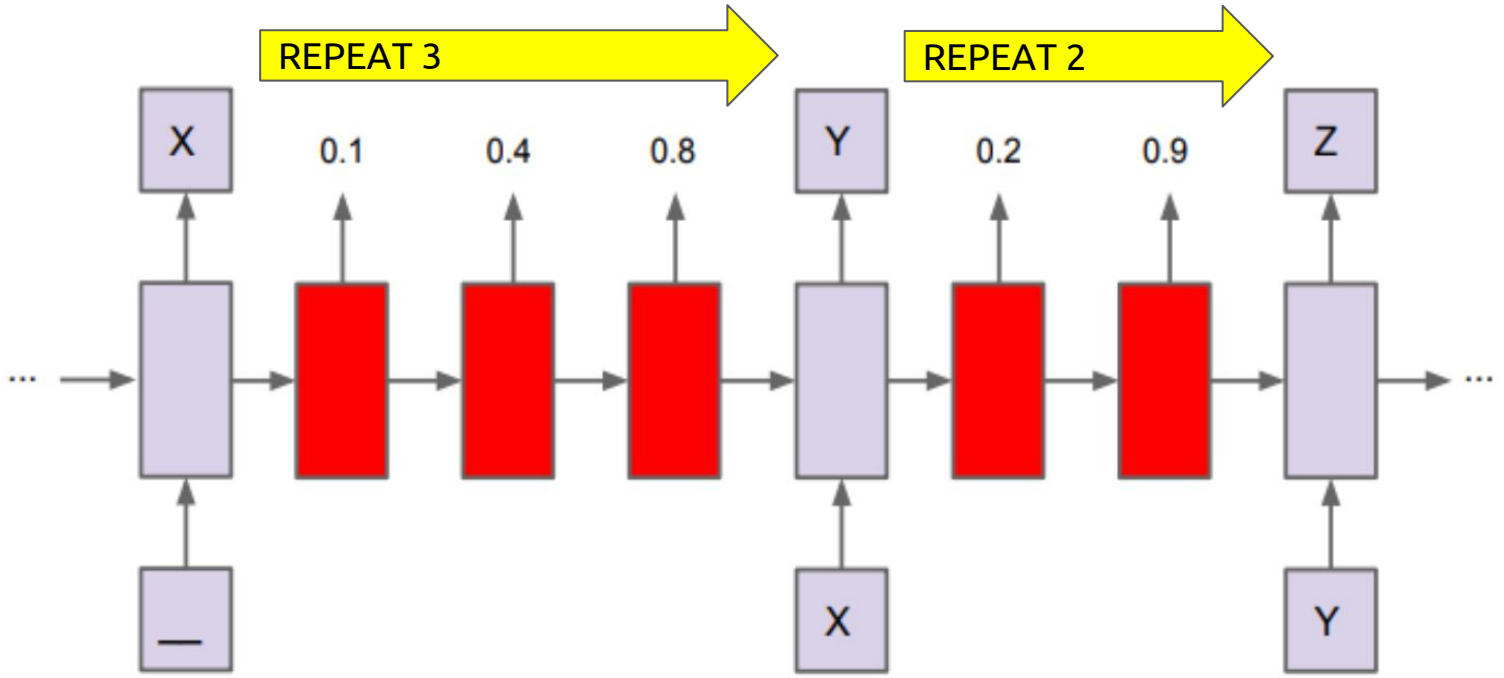


Víctor Campos

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.

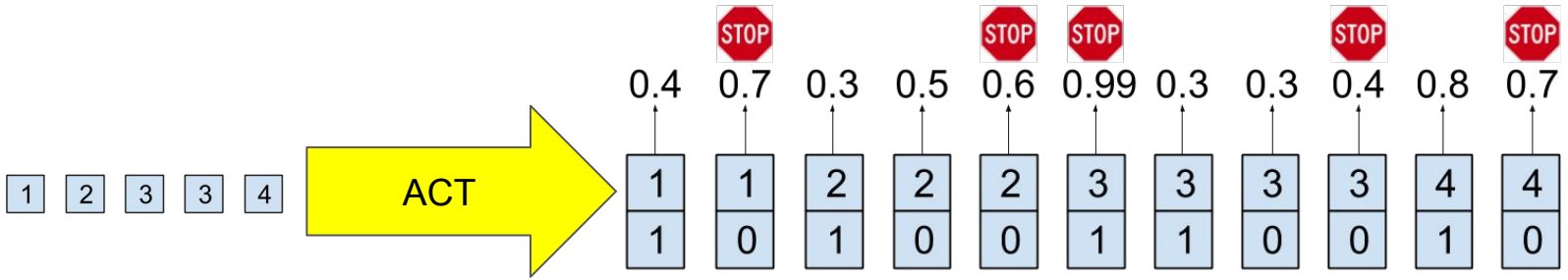
Adaptive Computation Time (ACT)



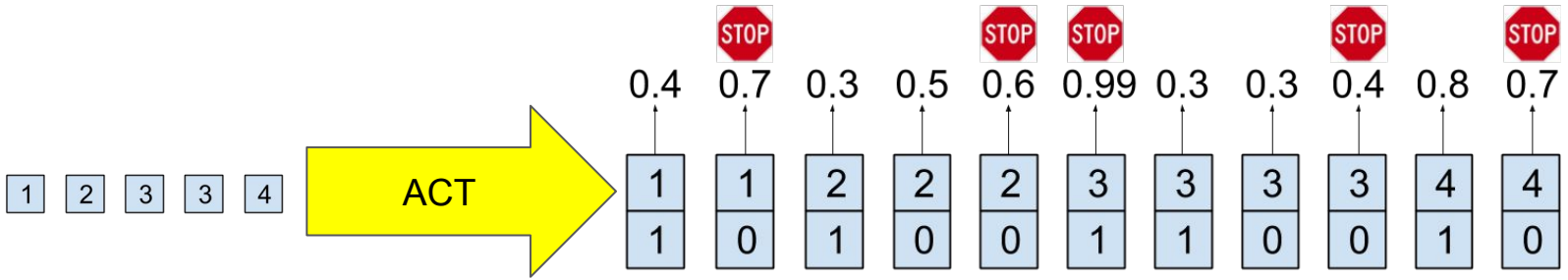
Google DeepMind

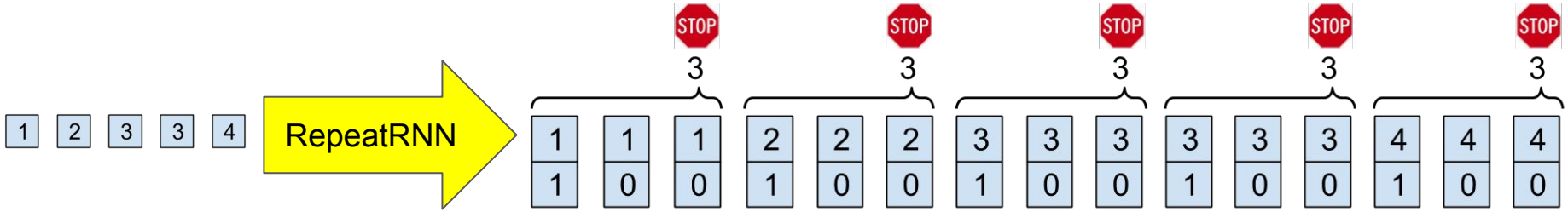
Graves, Alex. "Adaptive computation time for recurrent neural networks." arXiv preprint arXiv:1603.08983 (2016).

Adaptive Computation Time (ACT)

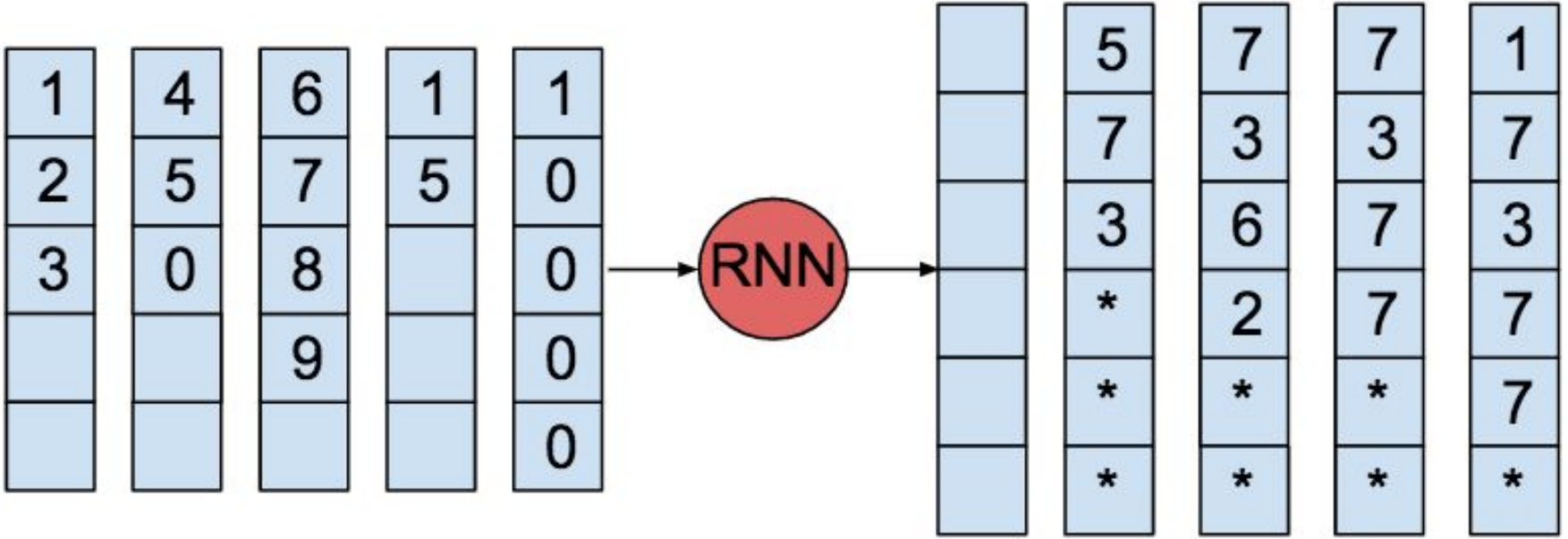


ACT vs RepeatRNN





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

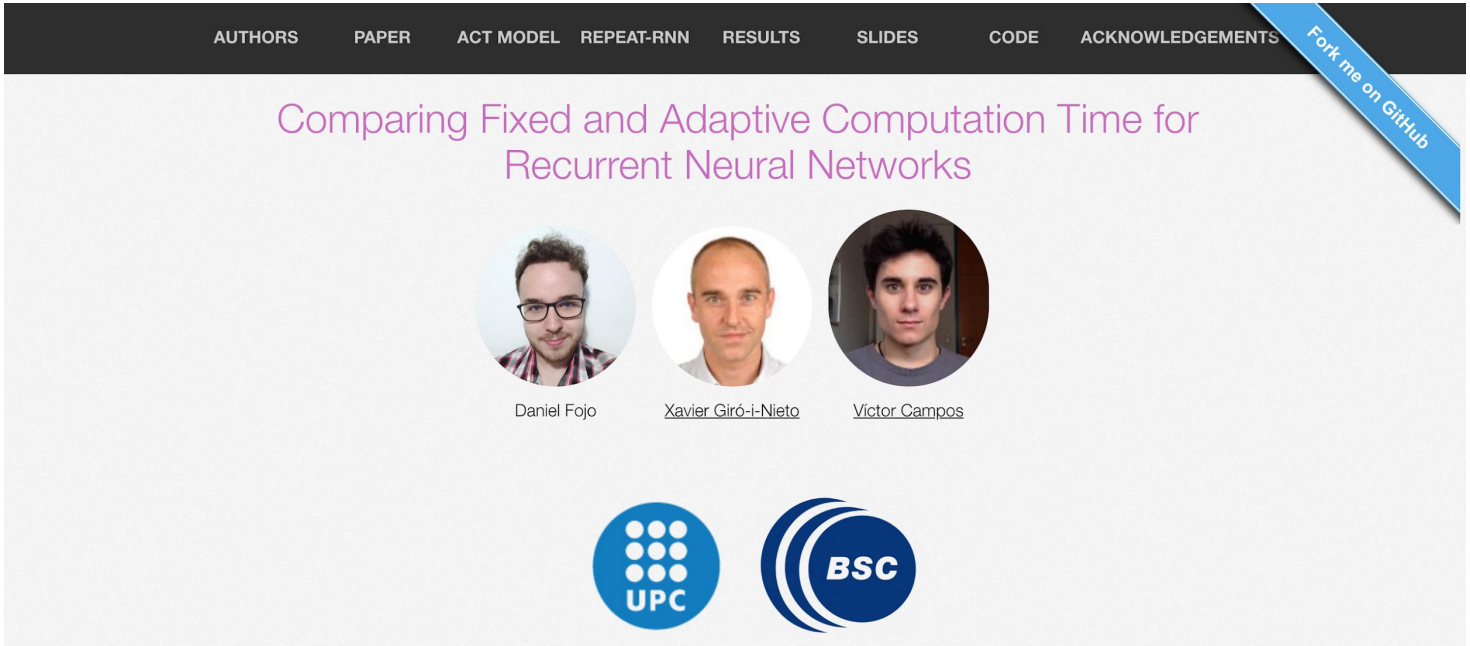
Faster training....

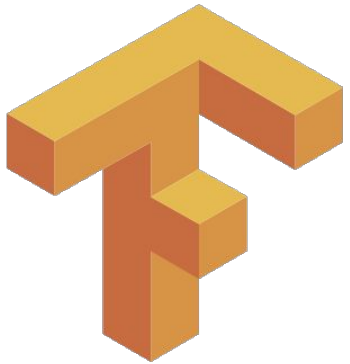
Experiment: One-hot addition

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





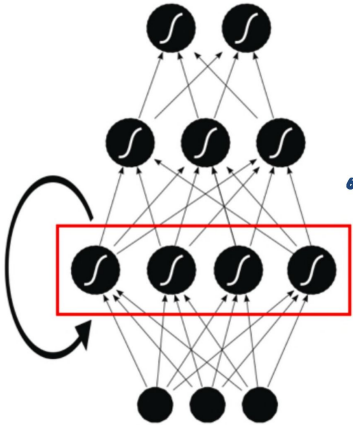
bit.ly/repeat-rnn

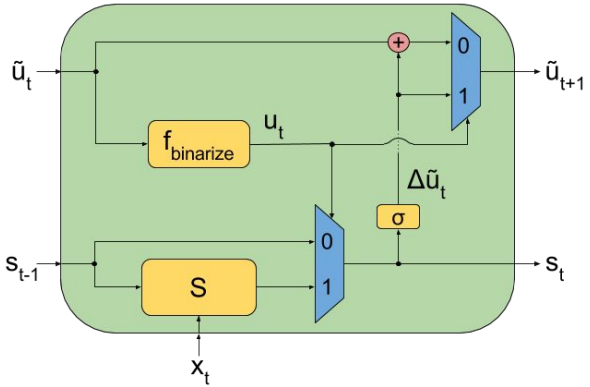
Outline

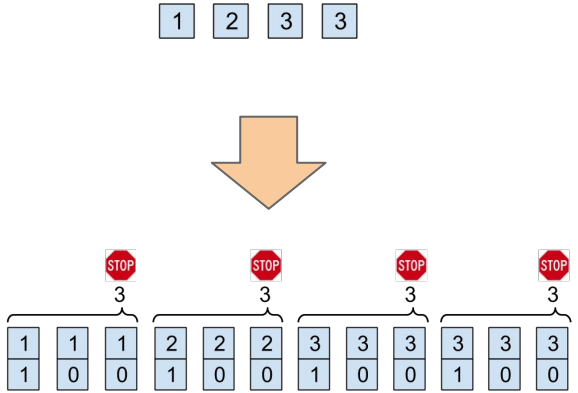
Recurrent Neural Networks (RNNs)

SkipRNN [ICLR 2018]

RepeatRNN [ICLRW 2018]







Acknowledgements

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

NVIDIA GPU CENTER OF EXCELLENCE

EXCELENCIA SEVERO OCHOA

Obra Social
Fundació "la Caixa"

Generalitat de Catalunya
Departament d'Economia i Coneixement

GOBIERNO DE ESPAÑA
MINISTERIO DE ECONOMÍA Y COMPETITIVIDAD

Deep Learning courses @ UPC (videos)

bit.ly/InsightDL2018
#InsightDL2018

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

Next edition Autumn 2018

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

Summer School (starts 28/06)

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

Next edition Winter/Spring 2019

Xavier Giro-i-Nieto

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

#InsightDL2018

Download slides from:
http://bit.ly/InsightDL2018

Click to comment:

We are looking for both industrial & academic partners:

xavier.giro@upc.edu

@DocXavi

Questions?