Enhanced Sequence Modeling with Deep Learning

Bálint Gyires-Tóth

BME TMIT

SmartLab
Intelligent Interactions

About us

SmartLab @ BME TMIT

~20 employees

Deep Learning Research

- Fundamental research (deep learning, reinforcement learning, HTM)
- Sequence and time series modeling
- Speech Technologies & NLP

Deep Learning Education

- Deep Learning project lab / BSc & MSc thesis / PhD
- Elective class: Deep learning in practice (next semester with 80 seats)
- NVidia Deep Learning Institute workshops and meetups

Deep Learning Trainings & Developments

About me

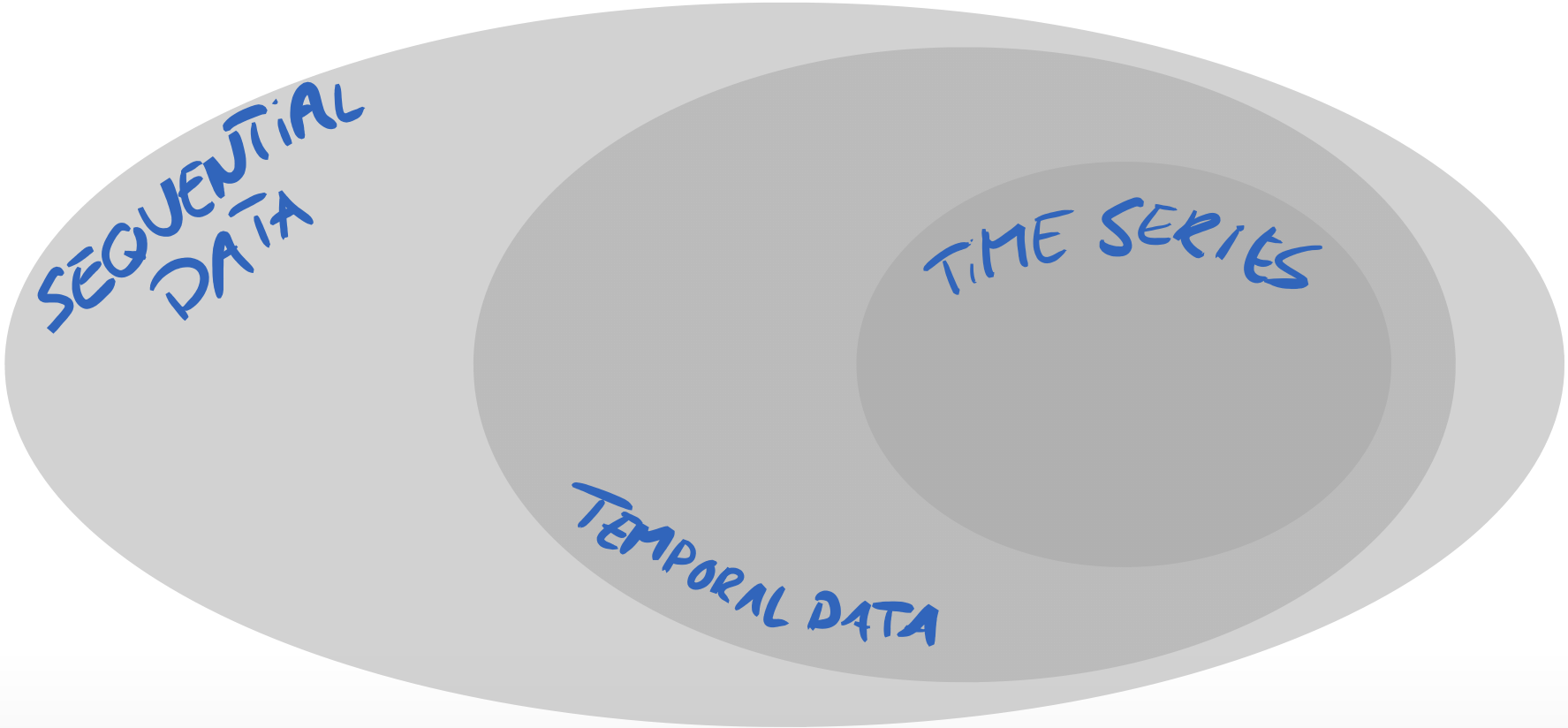
Bálint Gyires-Tóth, PhD

SmartLab @ BME TMIT

NVidia Deep Learning Institute
Certified Instructor &
University Ambassador

- Signal processing
- Time series modeling
- Machine and deep learning since 2007

Sequential data, temporal data, time series



Main challanges

Model correlations

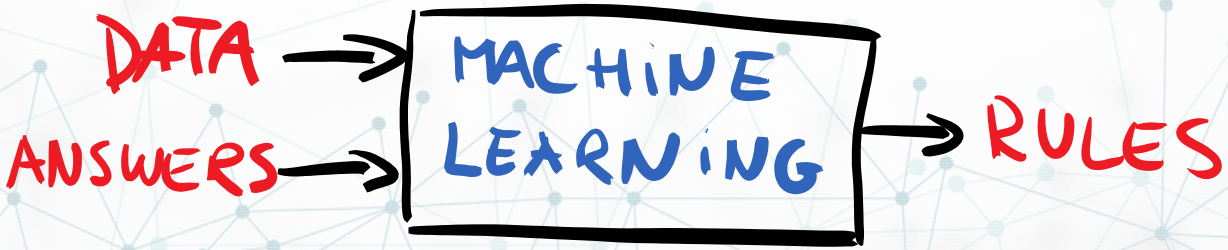
Model multiscale events

Increase the receptive field

Let’s model…

…sequential data.


DATA
RULES
CLASSICAL
PROGRAMING
ANSWERS


DATA
ANSWERS
MACHINE LEARNING
RULES

Deep learning is the key?

'Deep learning is part of a broader family of machine learning methods based on learning data representations.'

In practice: deep neural networks

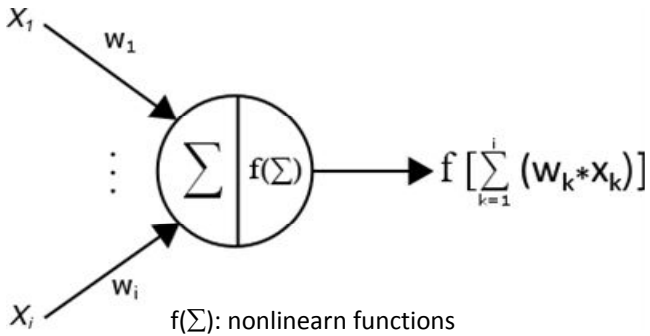
Whats new?

Algorithms + data + GPU

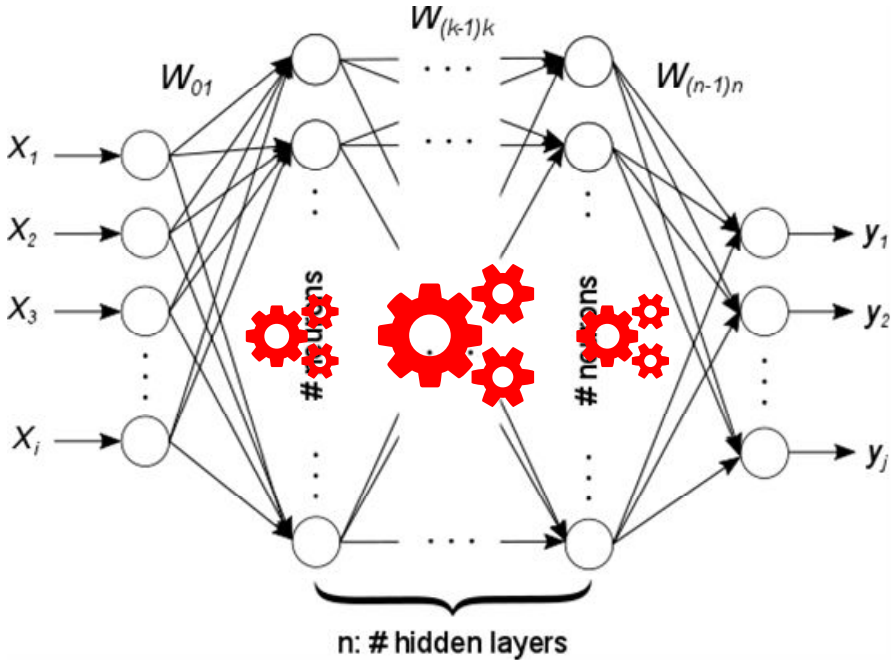
+ lower entry level

+ open-source research society („democratizing AI")

Feed forward neural networks



f(∑): nonlinearn functions
eg. f(∑)=max(0, ∑)

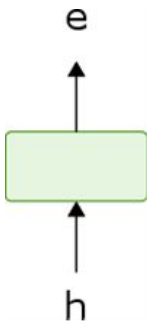


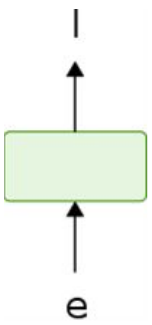
backpropagation

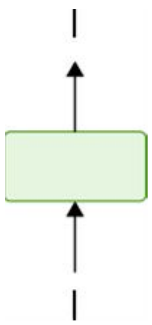
LEARNING: „fine-tuneing" weights

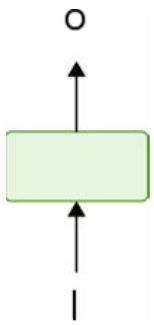
Context dependency in sequences

Feed Forward Neural Network

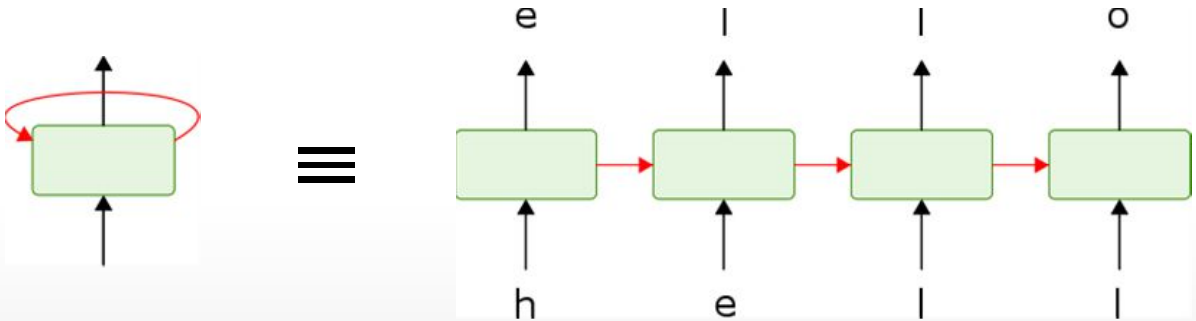
1. 

2. 

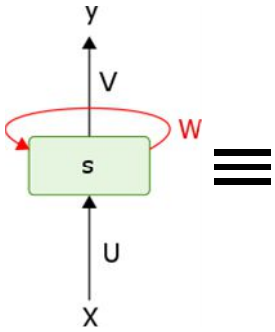
3. 

4. 

Recurrent Neural Network (RNN)

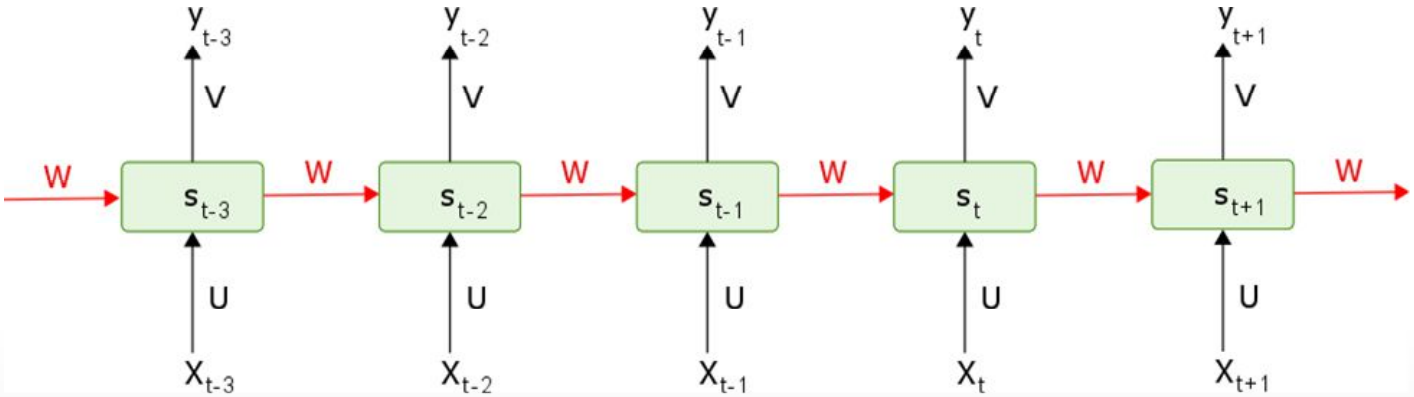


Unfolding RNNs



$$s_t = ReLU(Ux_t + Ws_{t-1}), \quad s_0 = \emptyset$$

$$y_t = softmax(Vs_t)$$



Problem?

Hard (impossible) to train

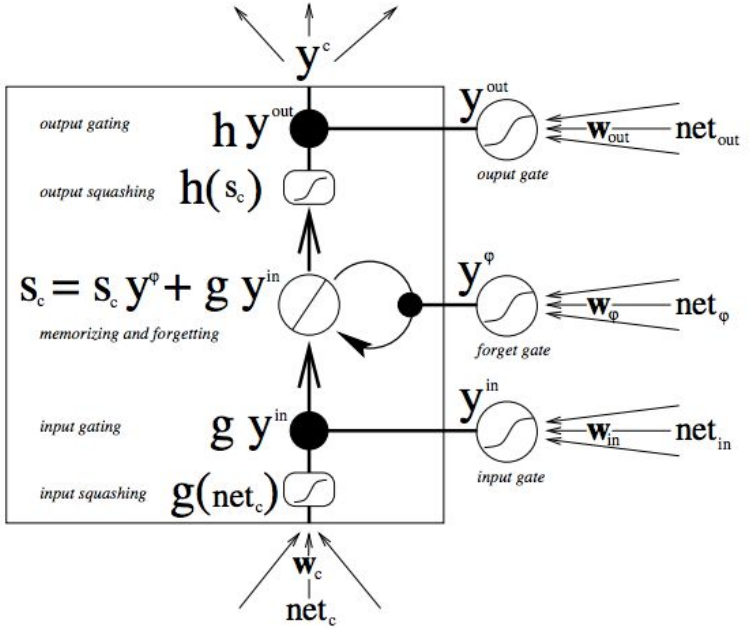
Receptive field: ~few tens

LSTM (Long Short-Term Memory) cell

Solves vanishing gradient problem.

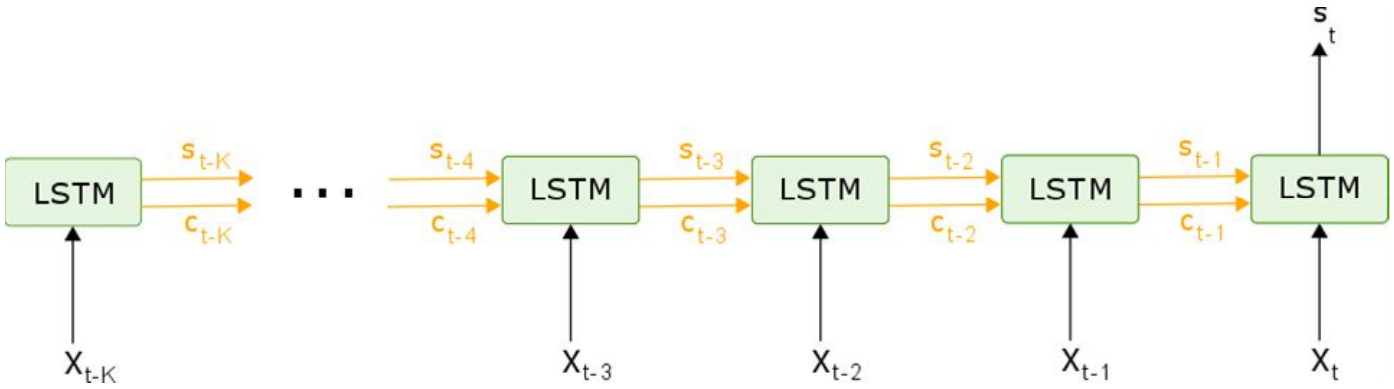
Gates:

- **Input gate:** decides which values to update
- **Forget gate:** decides to forget previous states and inputs
- **Output gate:** decides to keep the output for the next state or not
- Memory cell



Source: https://deeplearning4j.org/lstm.html

Unidirectional LSTM



ULSTM for predicting the next character

Orvosném pontosan elhallgatni **Szonyát**. Práfoldta a húgához illete valójában nem látta a lépcsőn, és felkapta fordult, és megfeledkeznék fogott:

- Mi a rábidult. **Hátrafordult, még mi?** - rikor intebbe, az is ugyanaznap tanú elotte, **Rogyion Romanovics**, hogy a múltért idemeted - szívelhoz! - mormogta mester, most kilette **Szvidrigajlov**? A nyitott a kapu. a kemenciortéhet, minden **jóformán ostobaság** - még harcolkában megint a szobájába, úgy általad, ha fogadja egy ugona zöld kajjának a **lépcsőházba**. Sokszor mit akarsz küldött, becsületes varrna, mint egy szavak.

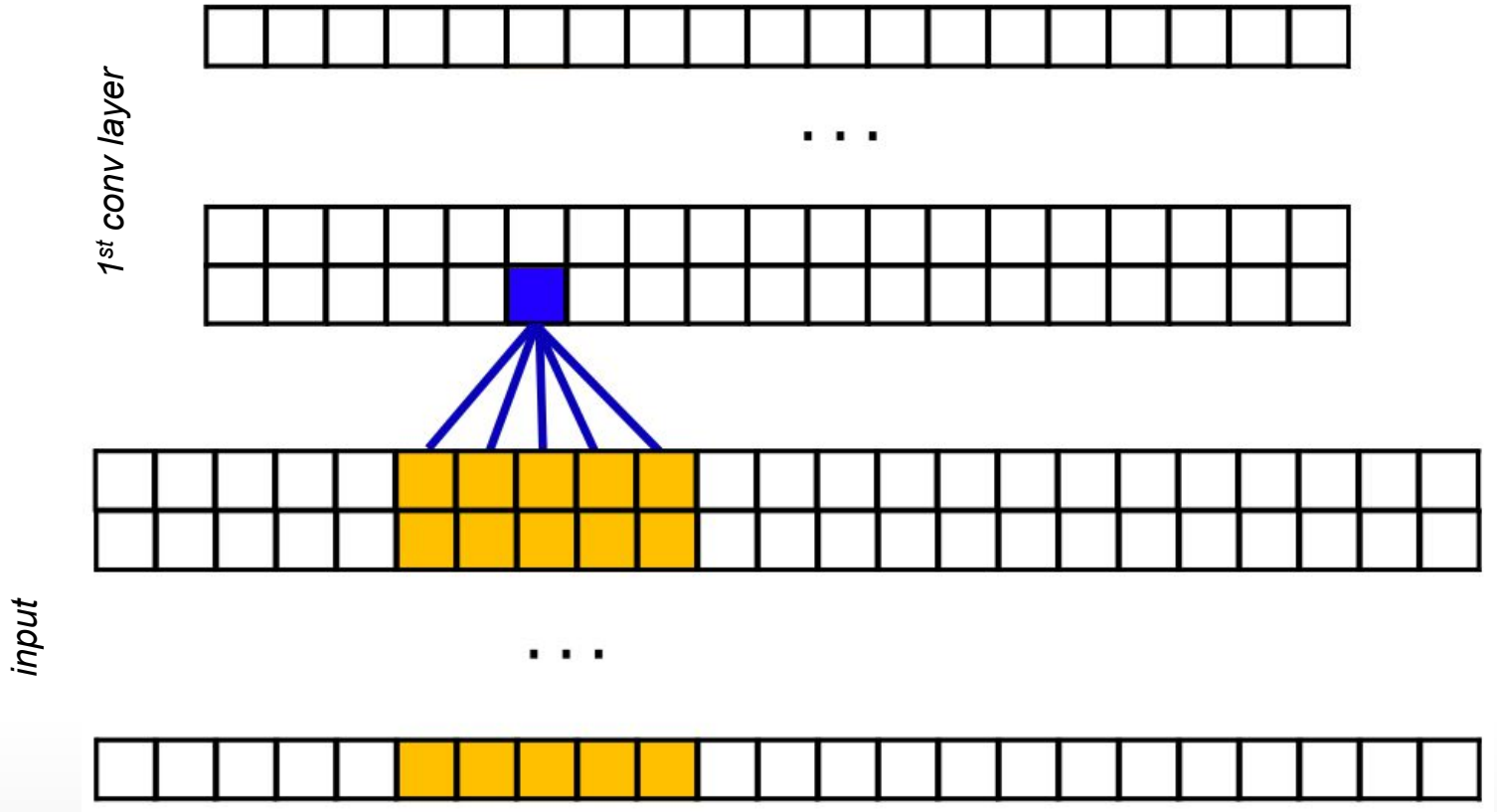
- Tettél! És most talán oda, szegény **remegés járta**, mennyire az lassan, **Lizaveta** menjenk az ilyen tényet hosszot... Jól meghelettem... Vehetek el hozzá? **Azt hiszi, kimerültség bolond**, jogos fején. Szonya jó socskát, alighanem remélem és lélegzett, annak vársz kínosan sokszor szorításukkal egy hónapja vagy tizenköteleben...

Receptive field: ~few tens..hundreds

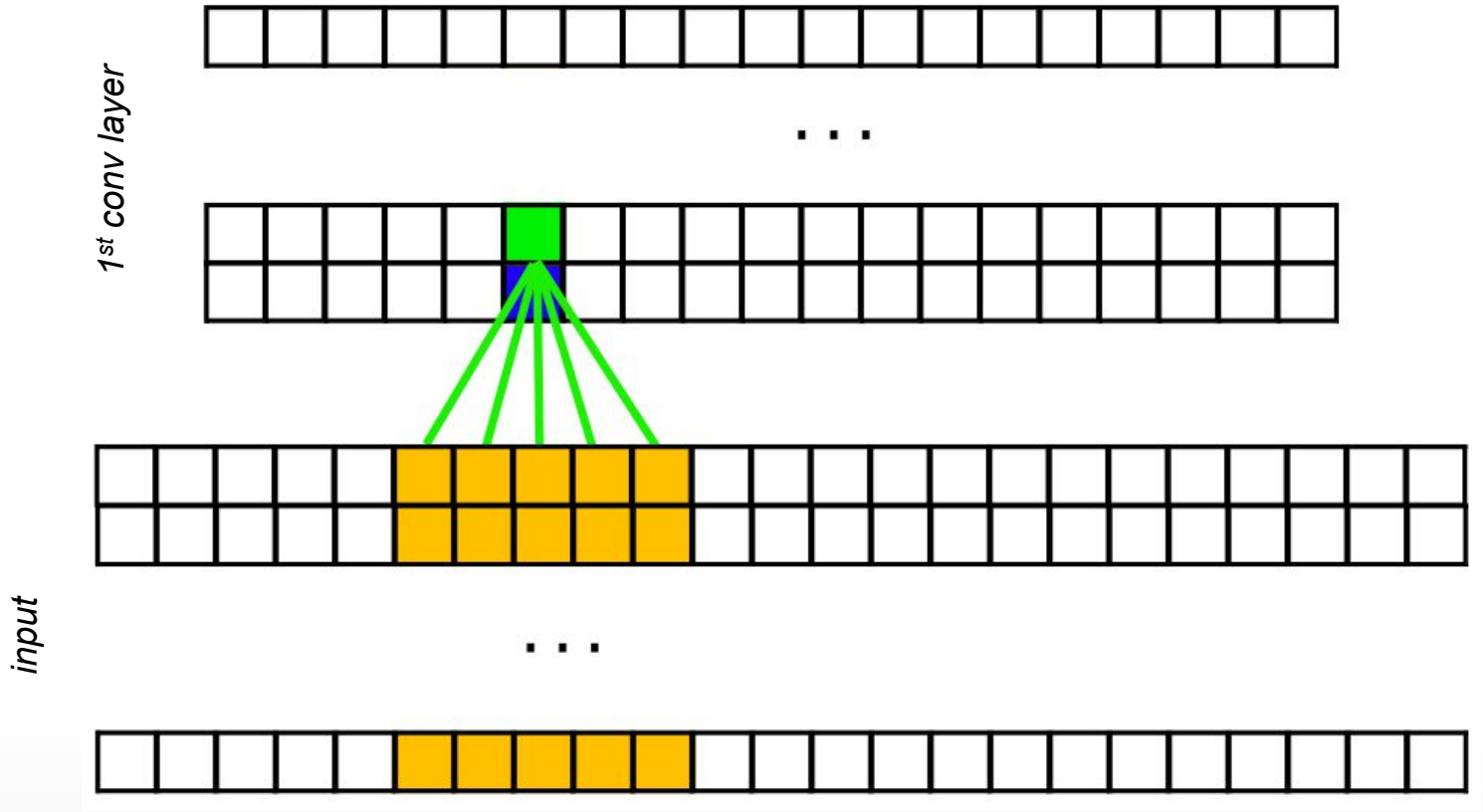
Problem?

Slow to train

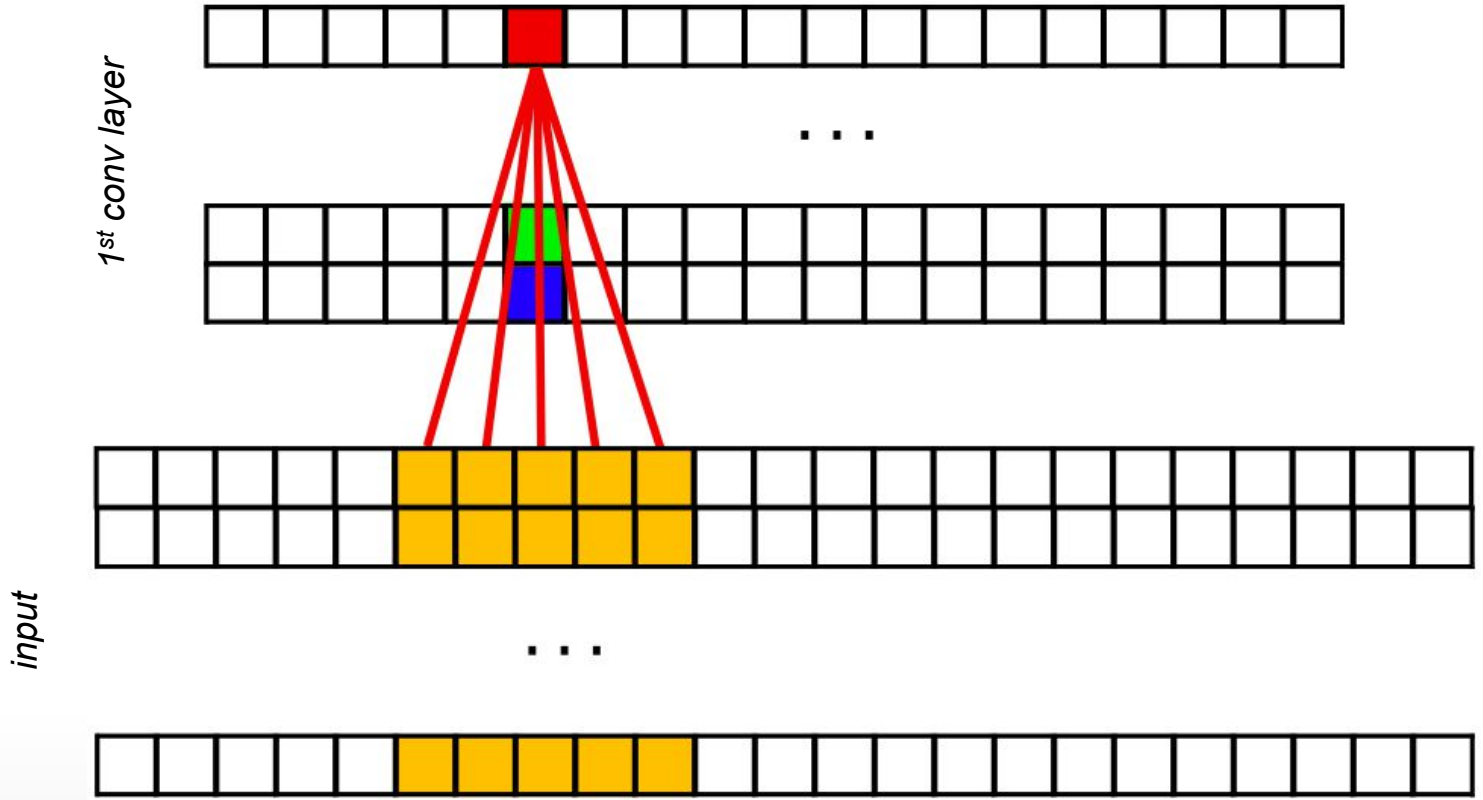
1D casual convolution



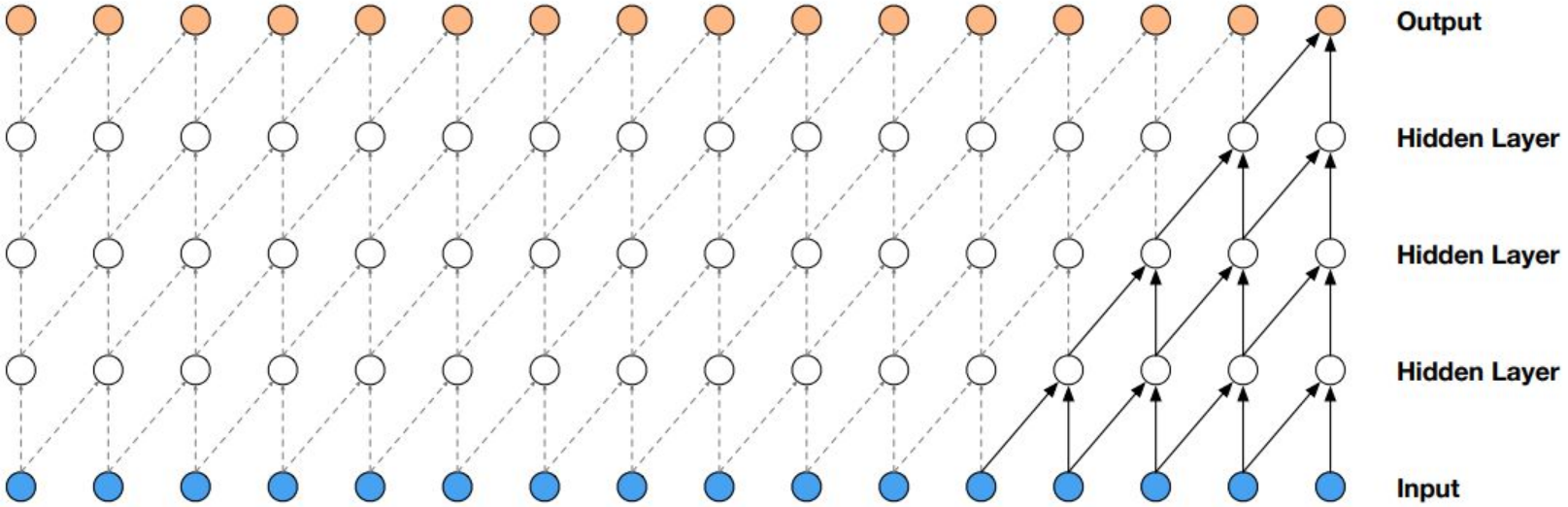
1D casual convolution



1D casual convolution



1D casual convolution



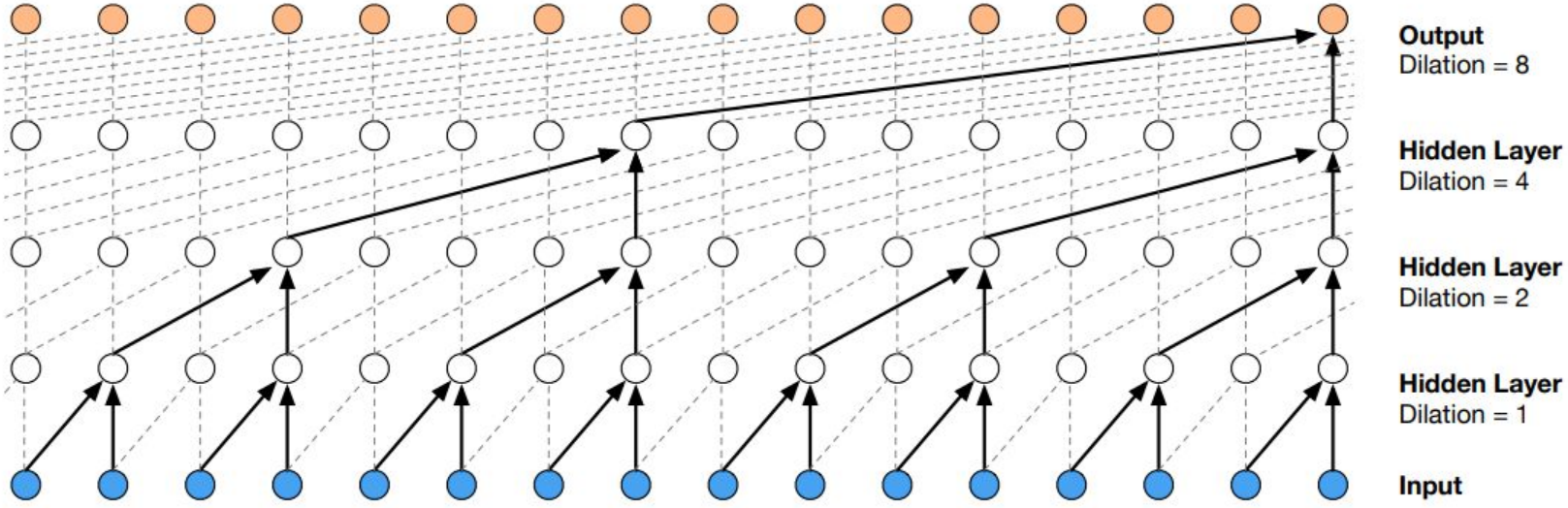
Source: Van Den Oord, Aaron, et al. "Wavenet: A generative model for raw audio." arXiv preprint arXiv:1609.03499 (2016).

Faster, receptive field: ~few tens..hundreds

Problem?

Large hiperparameter space

1D dilated convolution



Source: Van Den Oord, Aaron, et al. "Wavenet: A generative model for raw audio." arXiv preprint arXiv:1609.03499 (2016).

Receptive field: ~few thousands

Problem?

Large hiperparameter space

Large hiperparameter space

All we need is

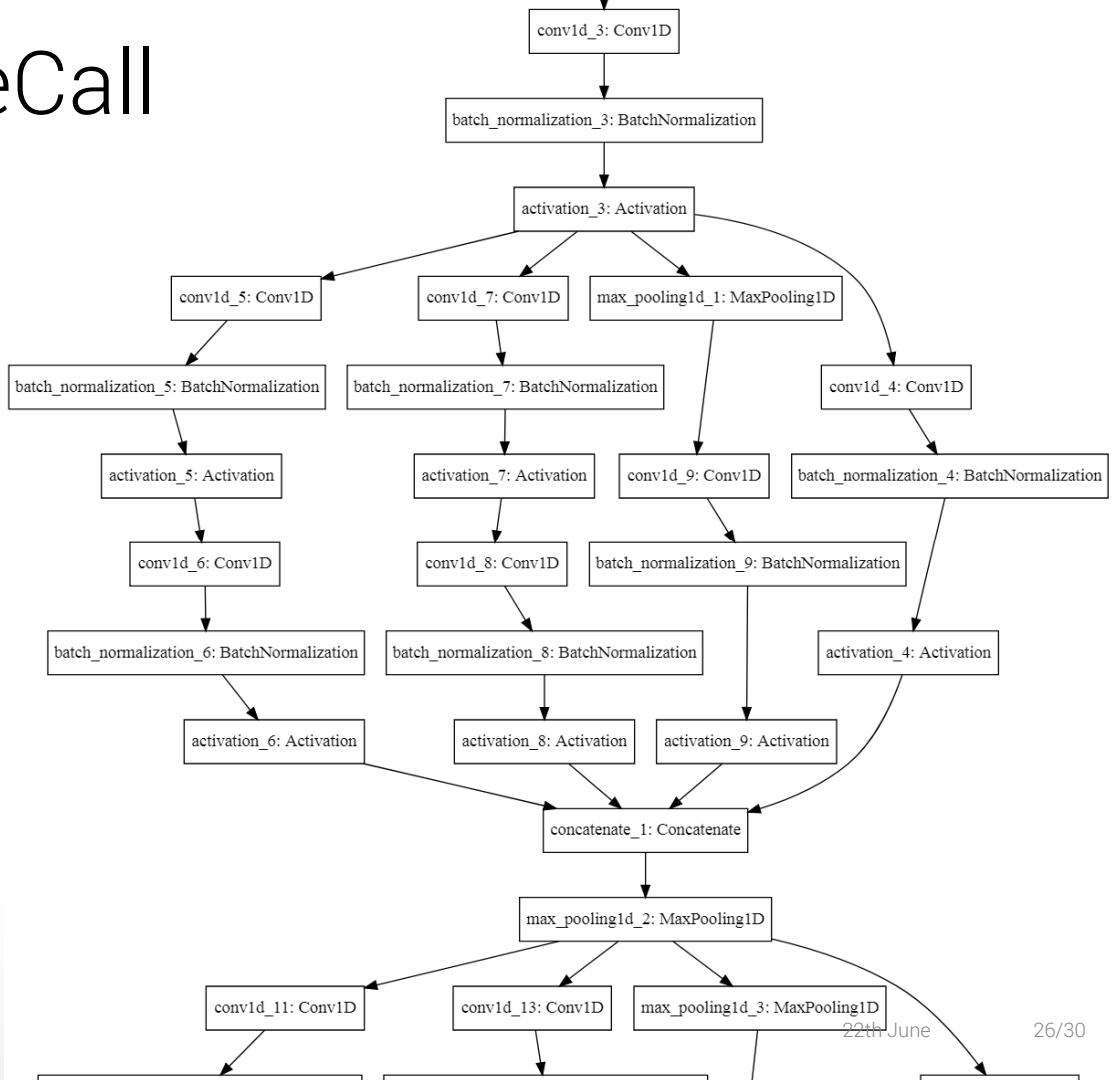
Application – SoleCall

INPUT:
200 x 7 sample points (1-2 seconds)

LAYERS (HYPEROPT):
16 residual blocks
filter (16+) and kernel size (3,5,7,9,11,…)

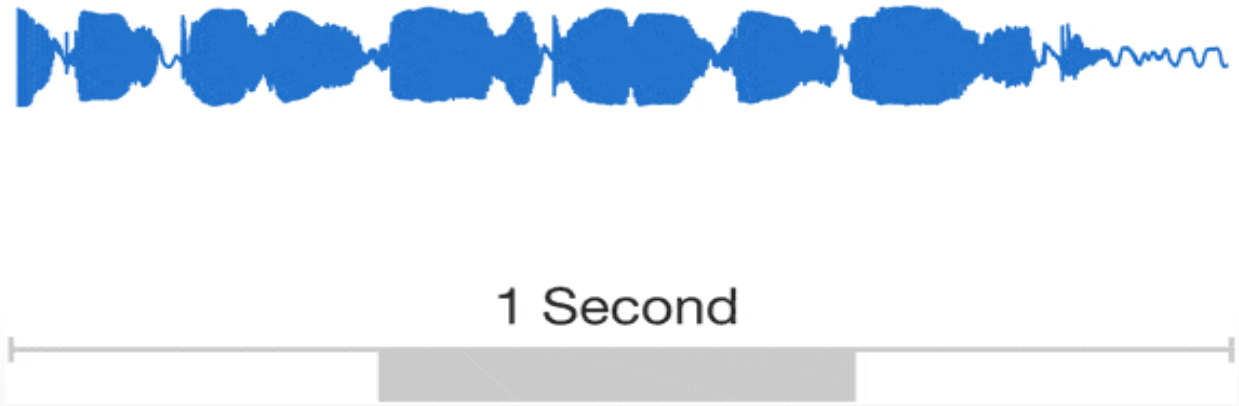
OUTPUT:
Binary classification

Over 99% accuracy.



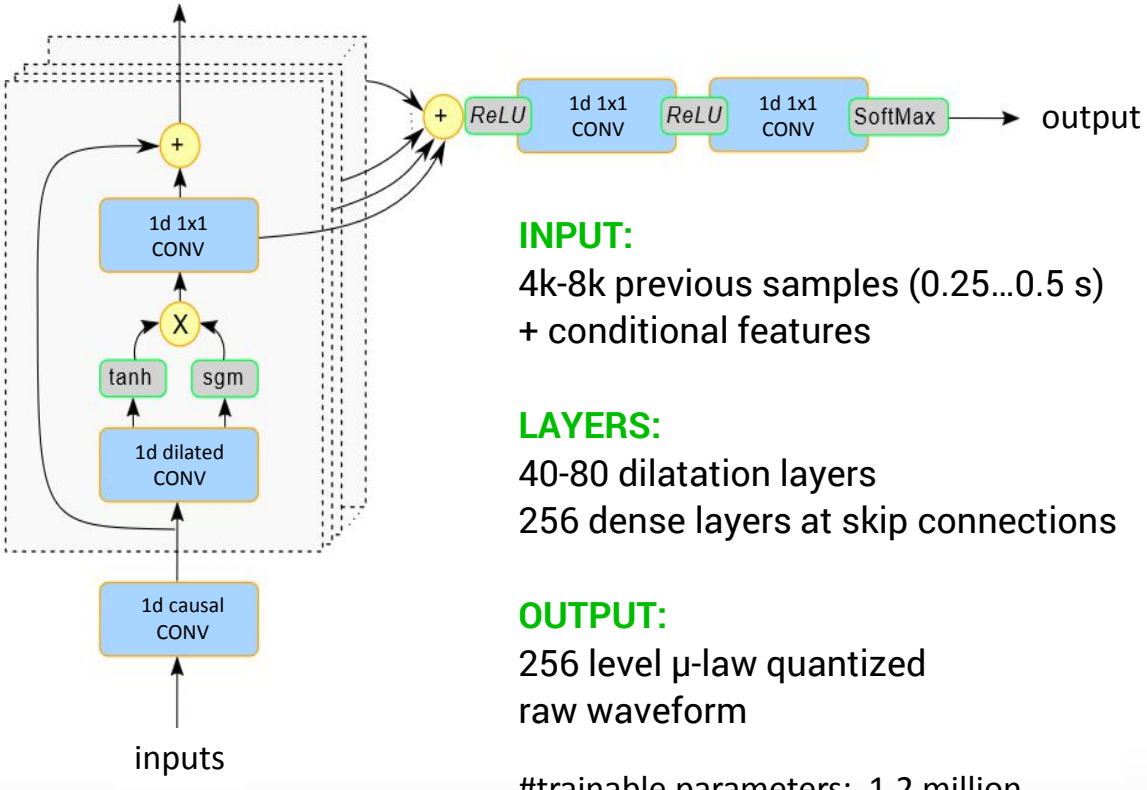
WaveNet

- Google DeepMind, Sept 2016 (later: Deep Voice, Baidu, February 2017)
- Idea comes from PixelCNN / PixelRNN
- Novel approach: model raw audio waveform



Source: https://deepmind.com/blog/wavenet-generative-model-raw-audio/

Application - WaveNet



(Based on DeepMind's WaveNet)

INPUT:
4k-8k previous samples (0.25…0.5 s)
+ conditional features

LAYERS:
40-80 dilatation layers
256 dense layers at skip connections

OUTPUT:
256 level μ-law quantized
raw waveform

#trainable parameters: 1.2 million

WaveNet samples

„A keleti tájakon kisebb eső is lehet.”

„Személyvonat indul Cegléd-Szolnok…”

„Feri bácsi megkapaszkodott, fújt egyet…”

Application scenarios

Google Duplex, call centers, railway stations, screen readers, mobile phones, automotive, virtual instruments, etc.

Summary

Deep learning for sequential data and time series.

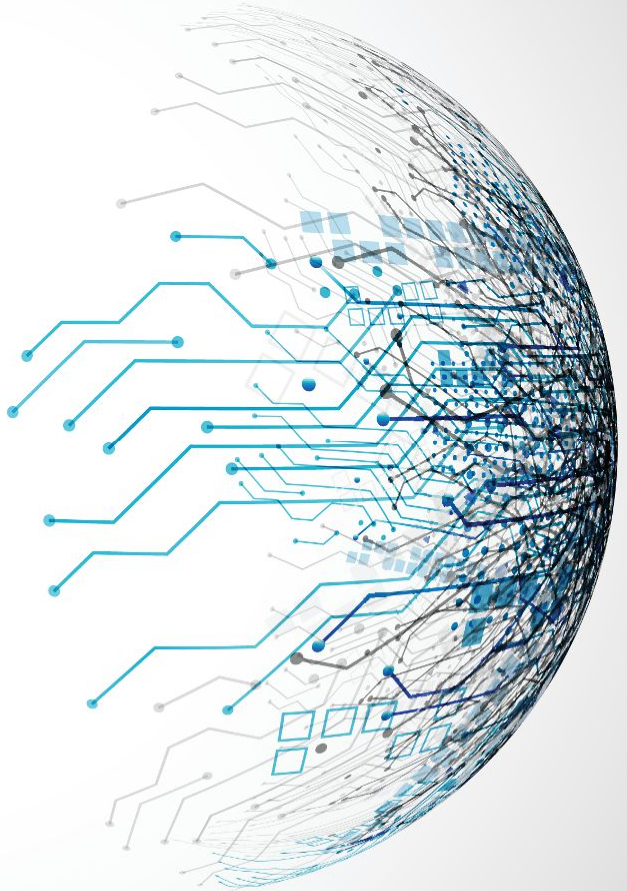
LSTM, Casual Convolution
Receptive field: ~few hundreds

Dilated Convolution
Receptive field: ~few thousands

Hyperopt still necessary

Thank you!

Bálint Gyires-Tóth
toth.b@tmit.bme.hu



BME TMIT

SmartLab
Intelligent Interactions

Some of the pictures are designed by Freepik.