Advanced eye tracking workshop

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Gaze event classifier

Georgii Zhulikov

26.10.2018

The problem

Your data:

Scene Frame Number	Scene Frame Time	Eye Frame Number	Eye Frame Time	Normal- ised Pupil_X	Normal- ised Pupil_Y	Normal- ised Gaze_X	Normal- ised Gaze_Y	Eye Movement Type	Patch Similarity
76.0	21314.81	85.0	21314.82	0.64064	0.58637	0.4498118	0.1327755	?	-0.985554
77.0	21314.85	86.0	21314.86	0.62303	0.61873	0.4862902	0.2143774	?	-0.985388
78.0	21314.88	87.0	21314.89	0.60714	0.63566	0.5181910	0.2608879	?	0.367969

The problem

Your data:

Scene Frame Number	Scene Frame Time	Eye Frame Number	Eye Frame Time	Normal- ised Pupil_X	Normal- ised Pupil_Y	Normal- ised Gaze_X	Normal- ised Gaze_Y	Eye Movement Type	Patch Similarity
76.0	21314.81	85.0	21314.82	0.64064	0.58637	0.4498118	0.1327755	?	-0.985554
77.0	21314.85	86.0	21314.86	0.62303	0.61873	0.4862902	0.2143774	?	-0.985388
78.0	21314.88	87.0	21314.89	0.60714	0.63566	0.5181910	0.2608879	?	0.367969

The solutions

1. Position based (dispersion threshold identification [I-DT], minimum spanning tree identification [I-MST])

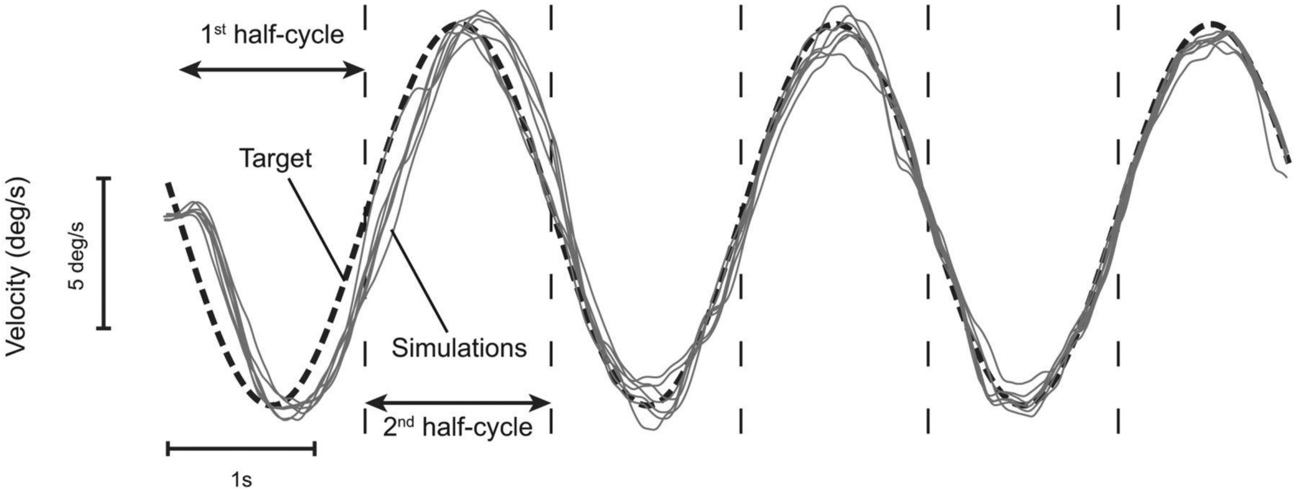
2. Velocity based (velocity threshold identification [I-VT], hidden Markov model identification [I-HMM], Kalman filter identification [I-KF])

3. Acceleration based (finite input response filter identification [I-FIR])

These work for “**fixation vs saccade**” classification.

Smooth pursuit

- Involuntary (can be produced only during experiments with dynamic scenes)
- Can't be detected as easily
  - Low velocity (like a fixation)
  - Change of position (like a saccade)



Solutions

- Algorithmic solutions
  - Agtzidis, et al. 2016 (clustering for several observers)
  - Larsson et al. 2015 (segmentation of intersaccadic intervals)
  - Dorr et al. 2010 (thresholding, not the aim of the study)
- Manual annotation
  - Has additional problems: variability, labor
  - Generally still better than algorithmic solutions

Solutions

- Algorithmic solutions
  - Agtzidis, et al. 2016 (clustering for several observers)
  - Larsson et al. 2015 (segmentation of intersaccadic intervals)
  - Dorr et al. 2010 (thresholding, not the aim of the study)
- Manual annotation
  - Has additional problems: variability, labor
  - Generally still better than algorithmic solutions **(or is it?)**

1D CNN-BLSTM

- One-dimensional (a linear sequence of eye movements)
- Convolutional neural network
    - A popular architecture primary used for finding patterns and features in data
    - Extracts data features into an abstract representation
- Bidirectional (uses both past data and future data for a particular point in time)
- Long short-term memory
    - A popular architecture for sequence analysis
    - Extracts sequence features
- Introduced in a paper titled *‘Deep Learning vs. Manual Annotation of Eye Movements’*

1D CNN-BLSTM

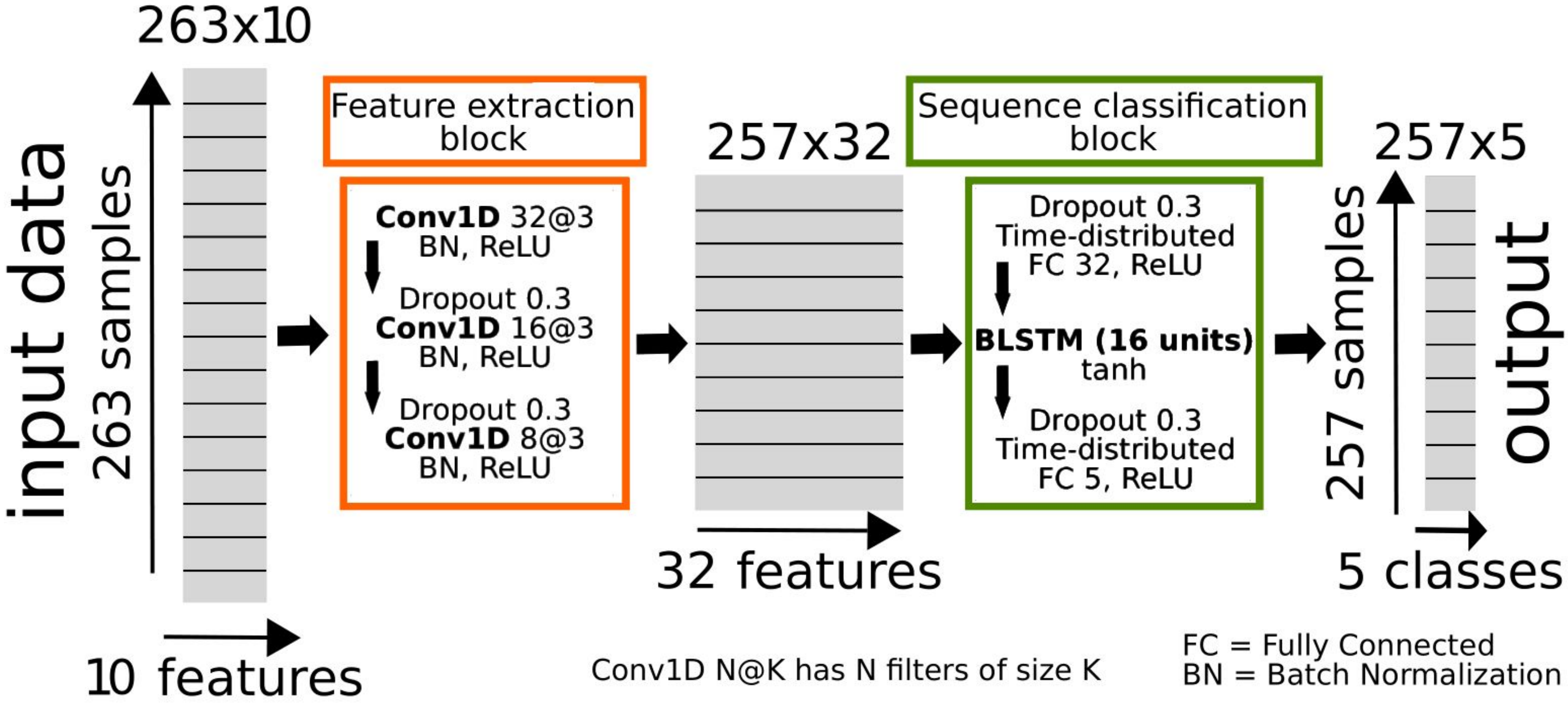
- How does it work?
- How does it use data?
- What are its limits and conditions?
- How well does it perform compared to other models?
- How well does it perform compared to humans?

Results

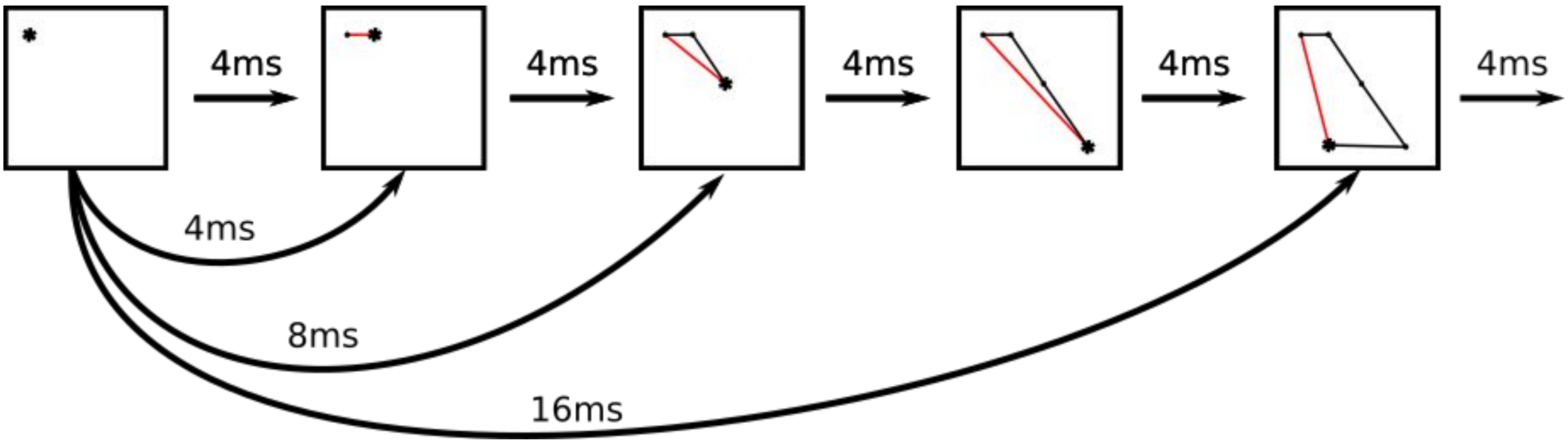
Model	Fixation F1	Saccade F1	Smooth Pursuit F1
1D CNN-BLSTM	**0.939**	**0.893**	**0.703**
[Agtzidis et al. 2016]	0.886	0.864	0.646
[Larsson et al. 2015]	0.912	0.861	0.459
I-VMP	0.907	0.725	0.570
[Dorr et al. 2010]	0.919	0.829	0.381
[Berg et al. 2009]	0.883	0.697	0.422

Model	Fixation F1	Saccade F1	Smooth pursuit F1
1D CNN-BLSTM	0.917	**0.863**	0.853
Expert 04	**0.934**	0.768	**0.905**
[Agtzidis et al. 2016]	0.886	0.809	0.871
Expert 03	0.927	0.720	0.895
Expert 02	0.877	0.776	0.765
Expert 01	0.805	0.831	0.748
I-VMP	0.814	0.722	0.696
[Berg et al. 2009]	0.703	0.570	0.437
[Dorr et al. 2010]	0.757	0.794	0.089
[Larsson et al. 2015]	0.744	0.790	0.0

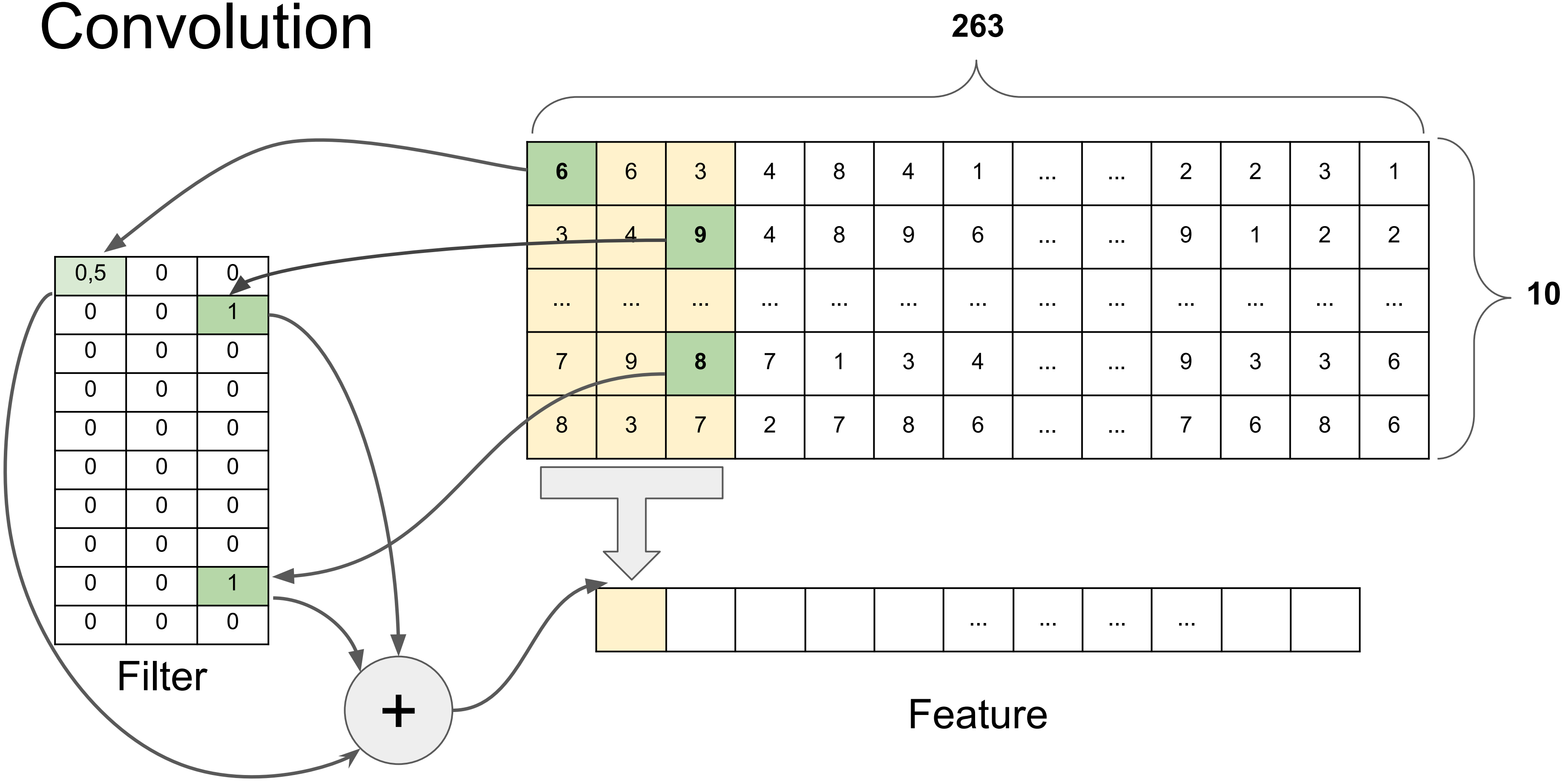
A look inside 1D CNN-BLSTM



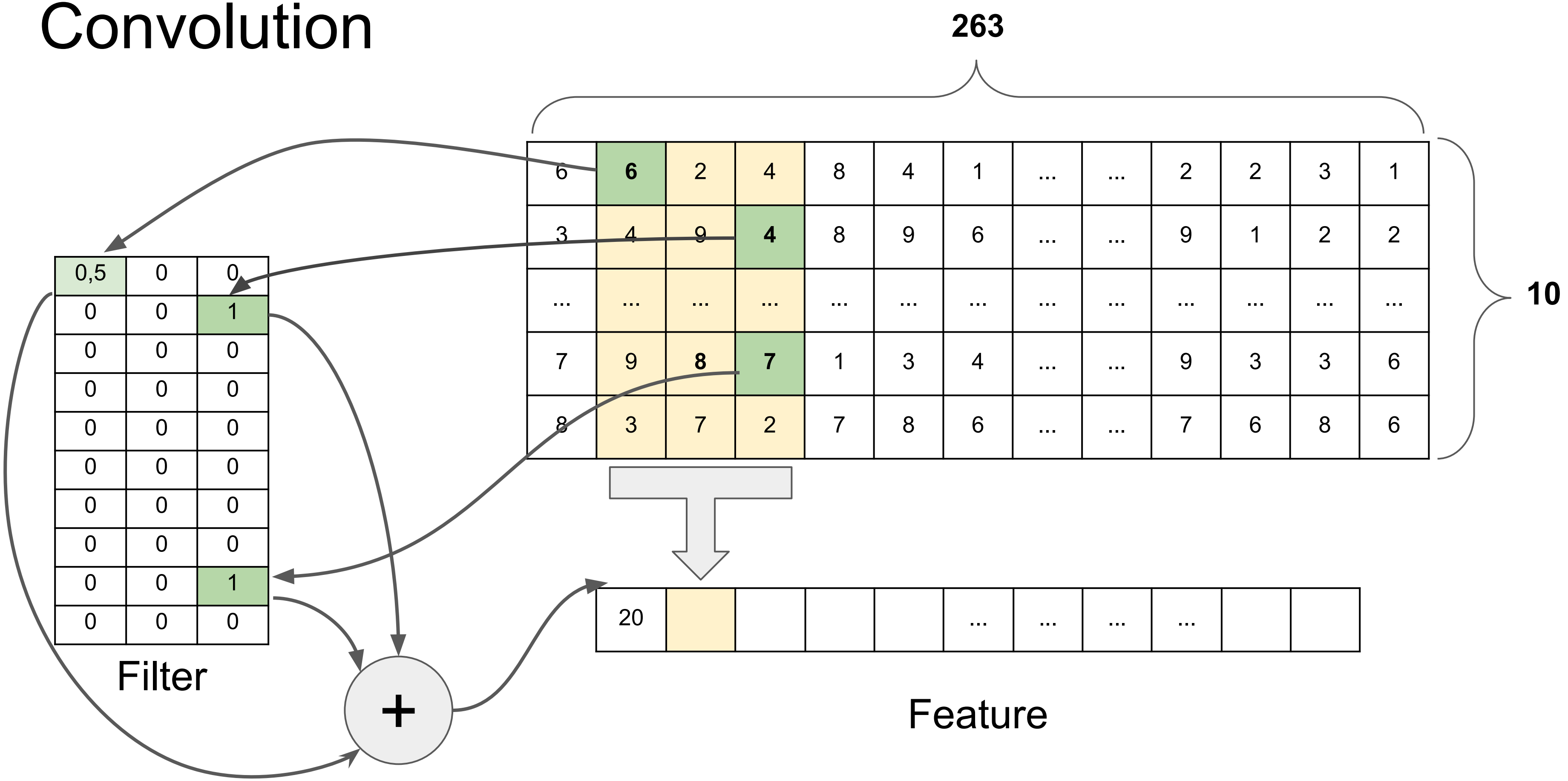
Input data



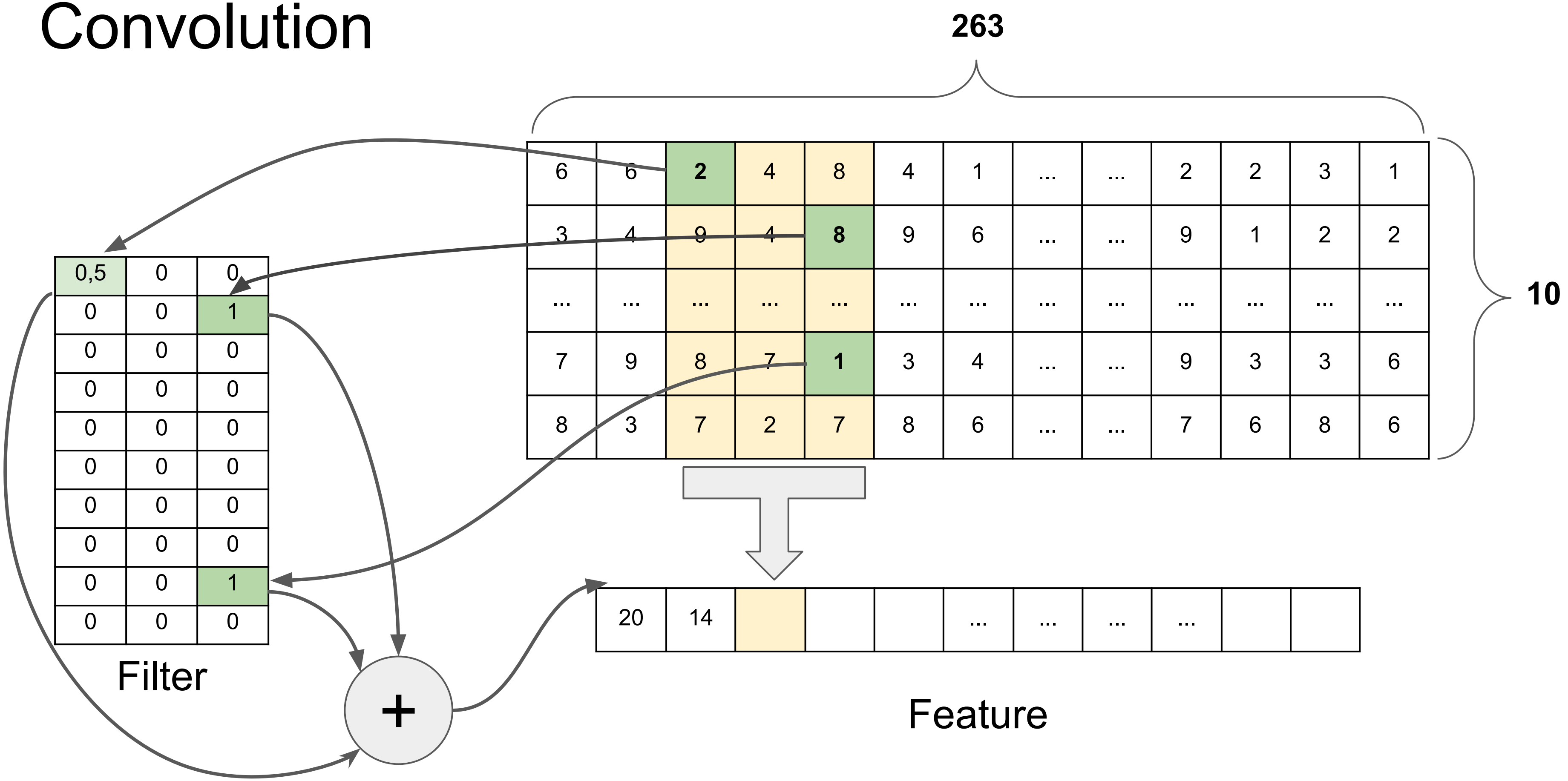
Convolution



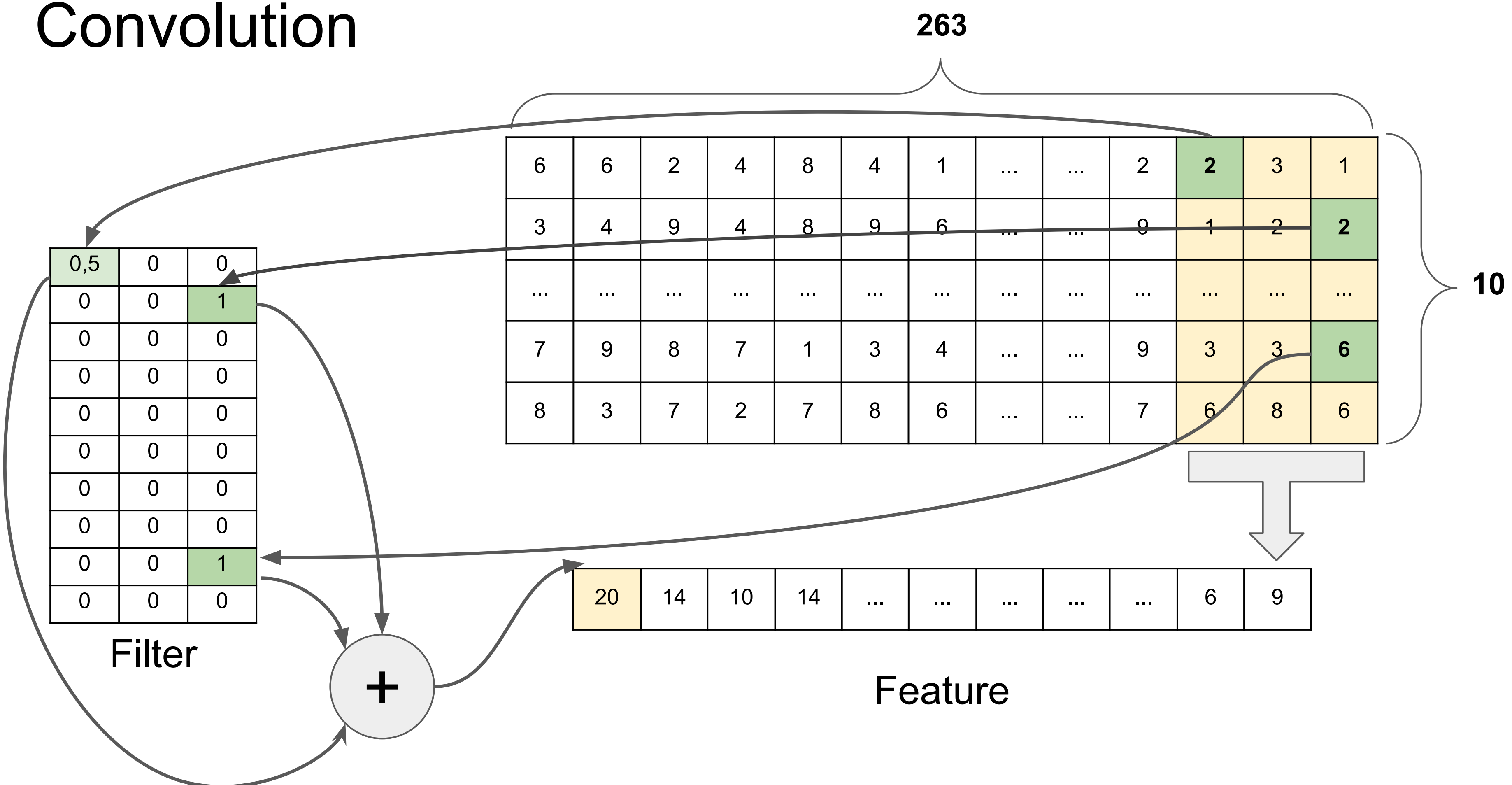
Convolution



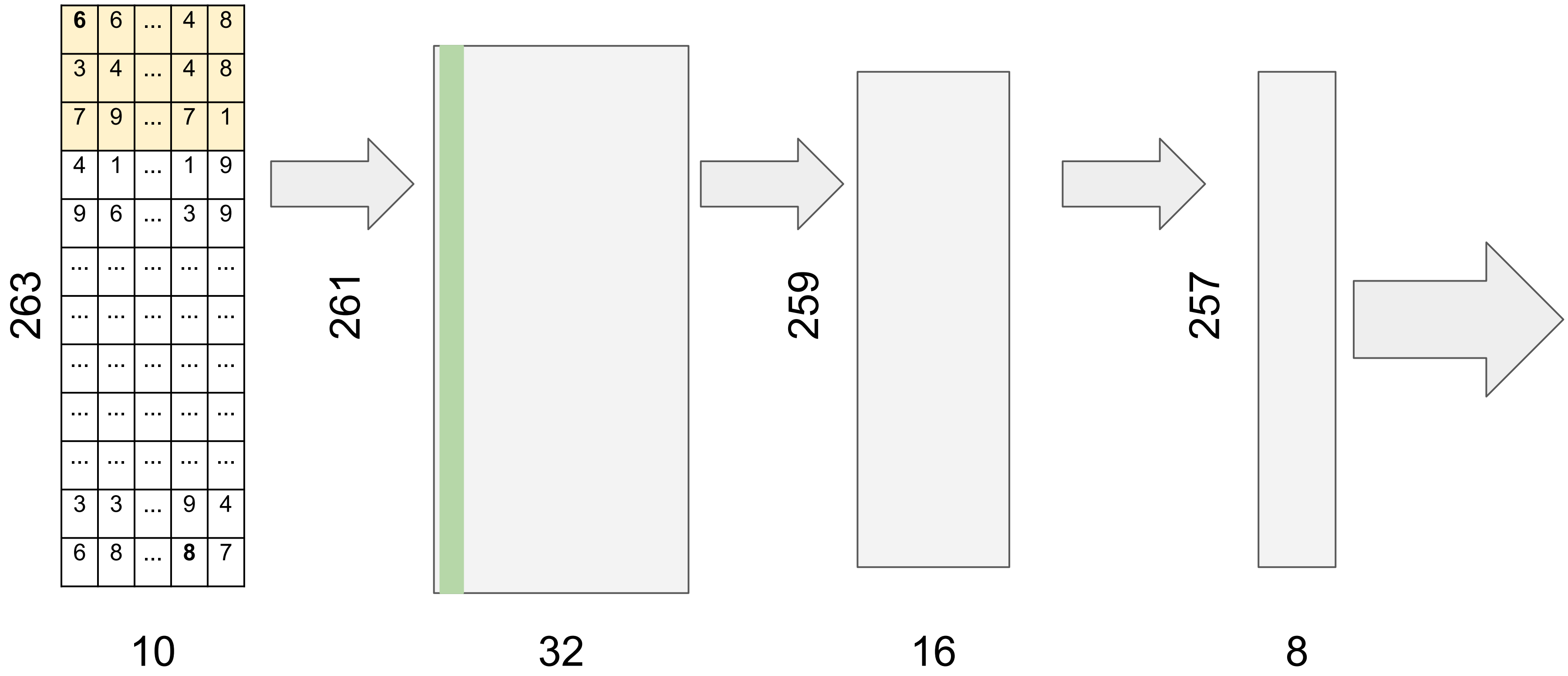
Convolution



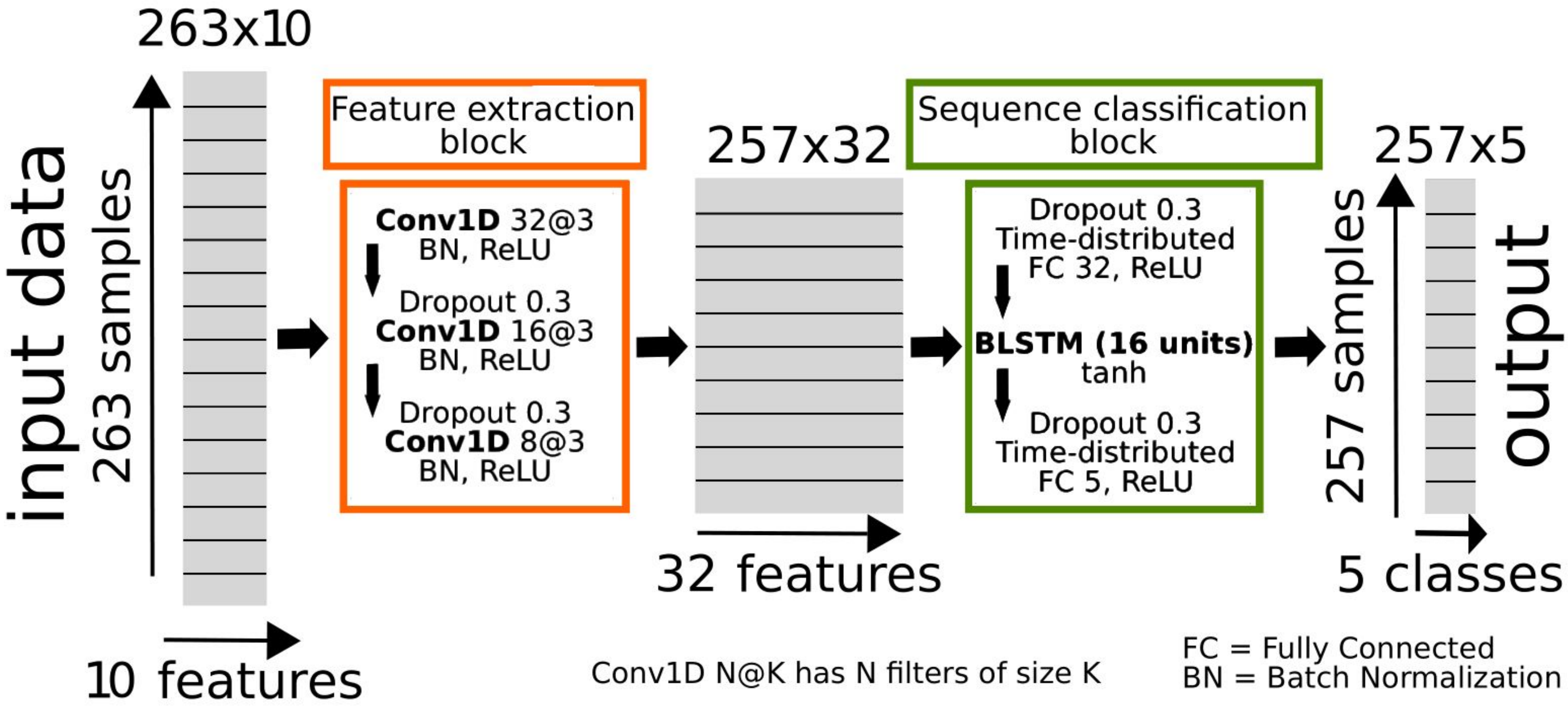
Convolution



Convolutional neural network

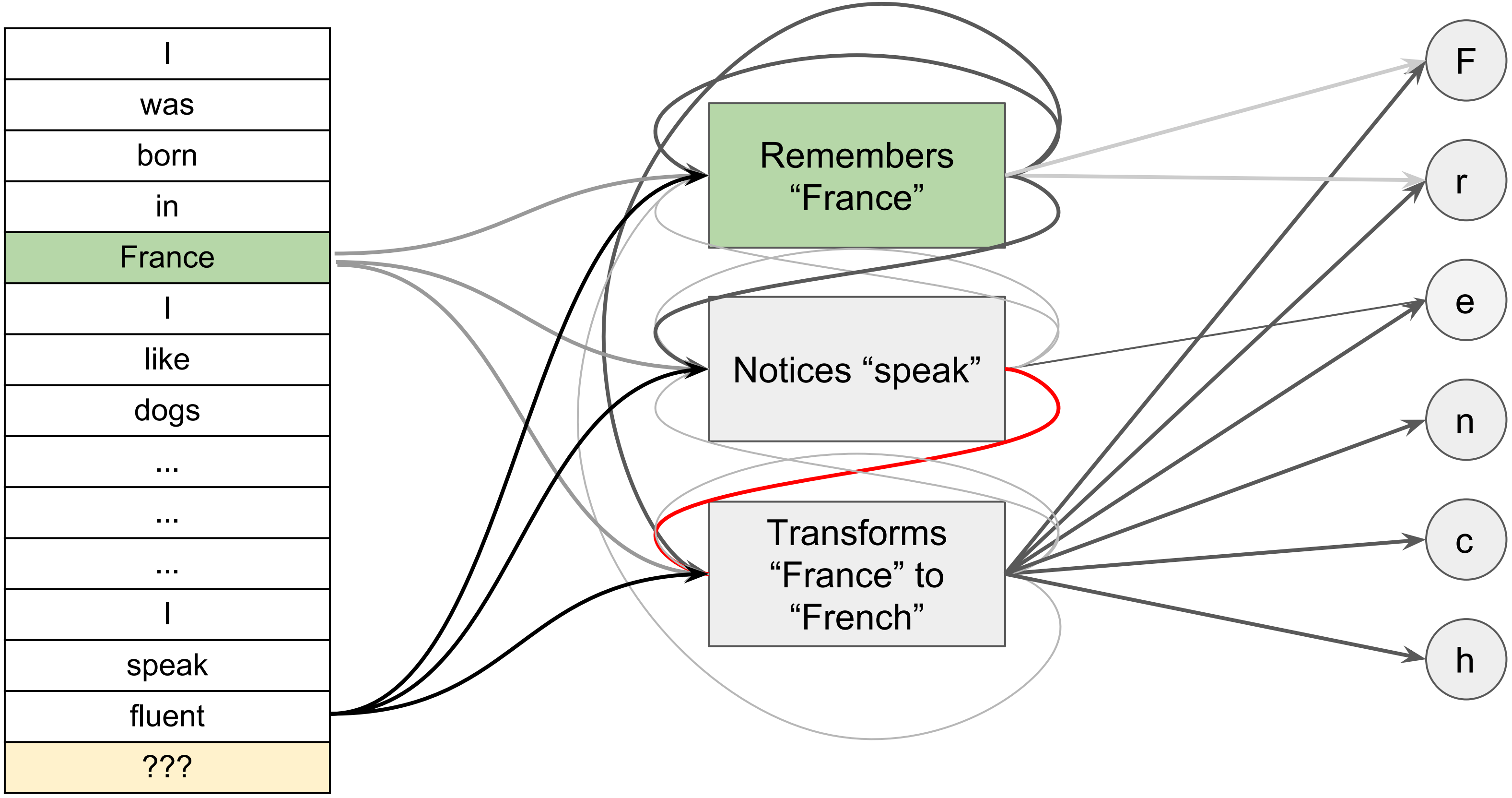


A remark



263x10
input data
263 samples
10 features
Feature extraction block
Conv1D 32@3
BN, ReLU
Dropout 0.3
Conv1D 16@3
BN, ReLU
Dropout 0.3
Conv1D 8@3
BN, ReLU
257x32
32 features
Sequence classification block
Dropout 0.3
Time-distributed
FC 32, ReLU
BLSTM (16 units)
tanh
Dropout 0.3
Time-distributed
FC 5, ReLU
257x5
257 samples
output
5 classes
Conv1D N@K has N filters of size K
FC = Fully Connected
BN = Batch Normalization

Analyzing sequences

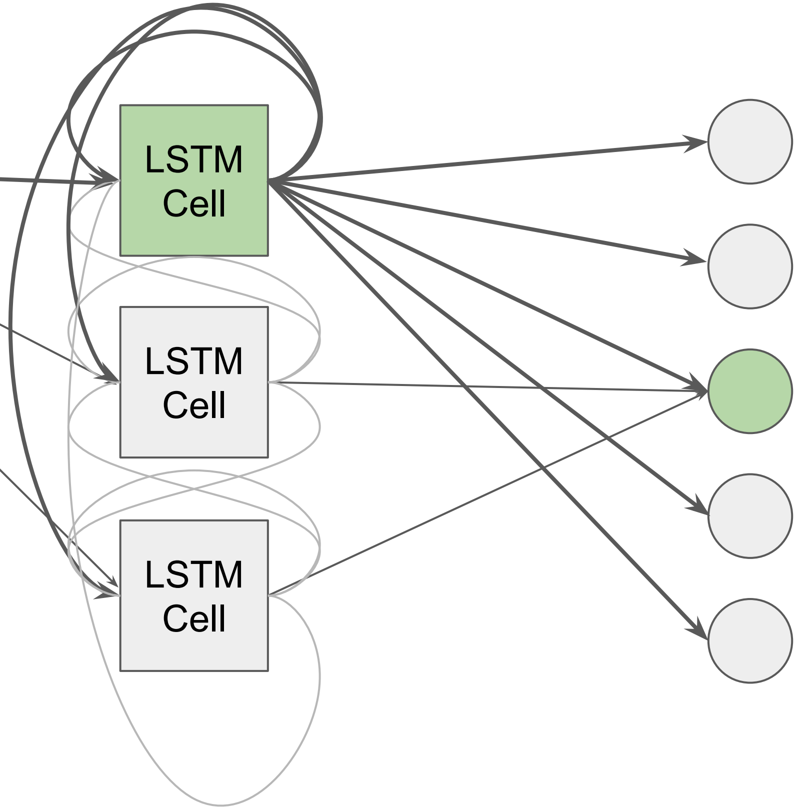


Bidirectional Long Short-Term Memory

257 elements

-26	...	-28	-8
1	...	5	-8
-19	...	-8	-10
21	...	40	6
-39	...	-8	15
...	...	...	...
...	...	...	...
...	...	...	...
15	...	24	-27
-9	...	38	-11
-26	...	7	-10

32 features



Results

Model	Fixation F1	Saccade F1	Smooth Pursuit F1
1D CNN-BLSTM	**0.939**	**0.893**	**0.703**
[Agtzidis et al. 2016]	0.886	0.864	0.646
[Larsson et al. 2015]	0.912	0.861	0.459
I-VMP	0.907	0.725	0.570
[Dorr et al. 2010]	0.919	0.829	0.381
[Berg et al. 2009]	0.883	0.697	0.422

Model	Fixation F1	Saccade F1	Smooth pursuit F1
1D CNN-BLSTM	0.917	**0.863**	0.853
Expert 04	**0.934**	0.768	**0.905**
[Agtzidis et al. 2016]	0.886	0.809	0.871
Expert 03	0.927	0.720	0.895
Expert 02	0.877	0.776	0.765
Expert 01	0.805	0.831	0.748
I-VMP	0.814	0.722	0.696
[Berg et al. 2009]	0.703	0.570	0.437
[Dorr et al. 2010]	0.757	0.794	0.089
[Larsson et al. 2015]	0.744	0.790	0.0

Discussion

- This classifier beats all other models at all types of eye movements
- This model beats humans at trying to mimic the human labeler
  - (just for saccades though)
- Ground truth data is hard to get, so more data will probably improve the model
- Modern machine learning approach proved to work well
  - More complex sequence analyzing algorithms could improve the result

Thank you

Convolutional neural network