ETTAC 2020

1st Workshop on Eye Tracking Techniques, Applications and Challenges

https://vision.unipv.it/ettac2020/

10 January 2021

In conjunction with

ICPR 2020

Eye Movement Classification with Temporal Convolutional Networks

Carlos Elmadjian
Candy Gonzales
Carlos H. Morimoto

University of São Paulo - Brazil

supported by

IME-USP

FAPESP
SÃO PAULO RESEARCH FOUNDATION

What is this work about?

What is this work about?

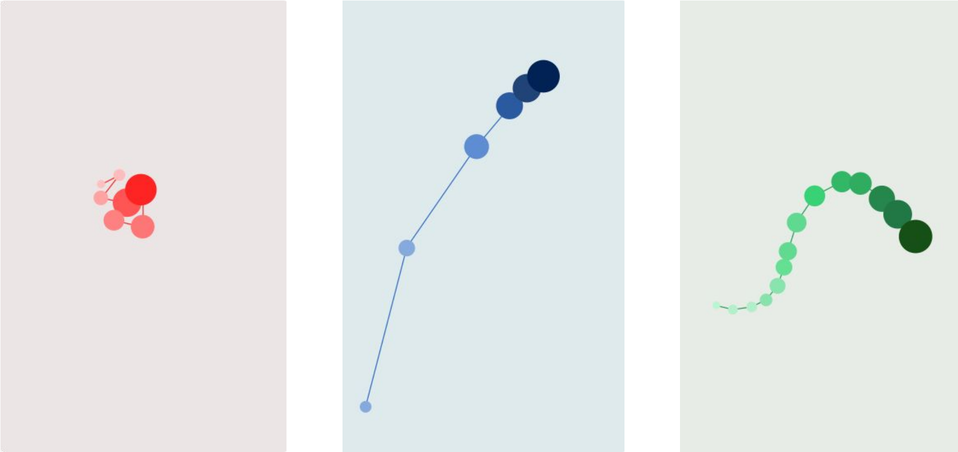
- improving the state of the art of the EMCP

What is this work about?

- improving the state of the art of the EMCP
- improving our understanding of the problem

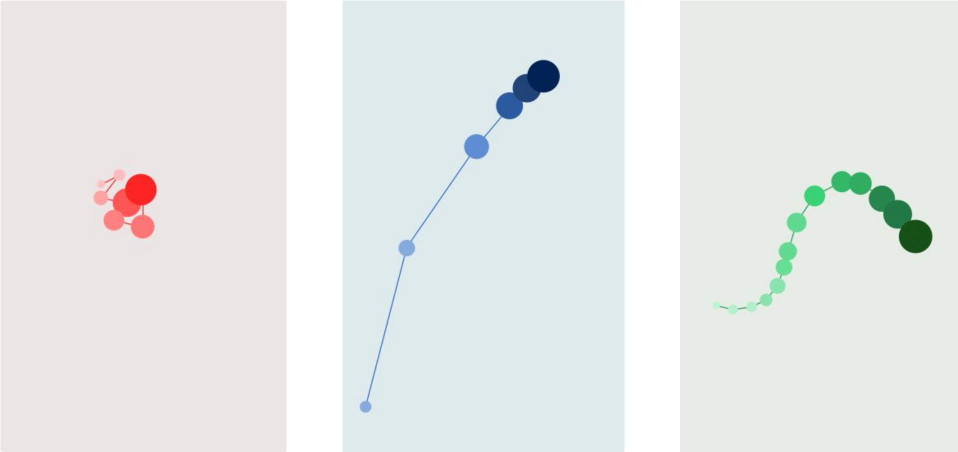
Problem statement: 3EMCP

Problem statement: 3EMCP



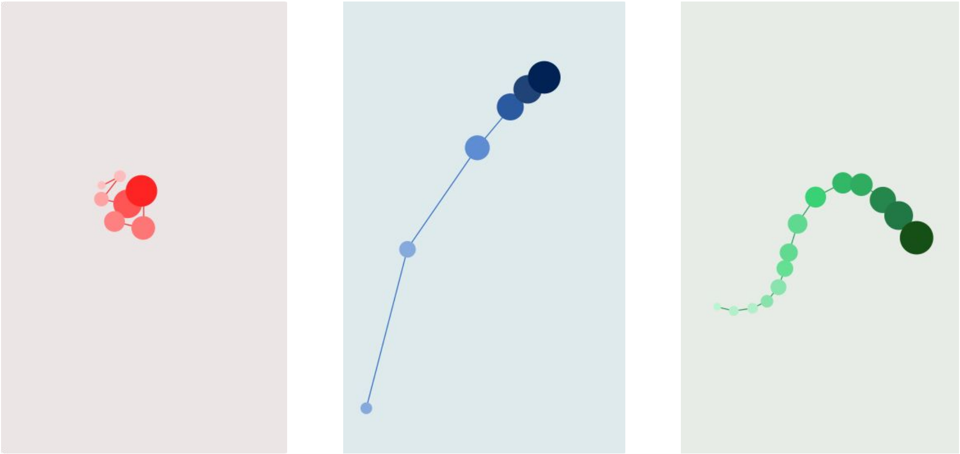
Fixations

Problem statement: 3EMCP



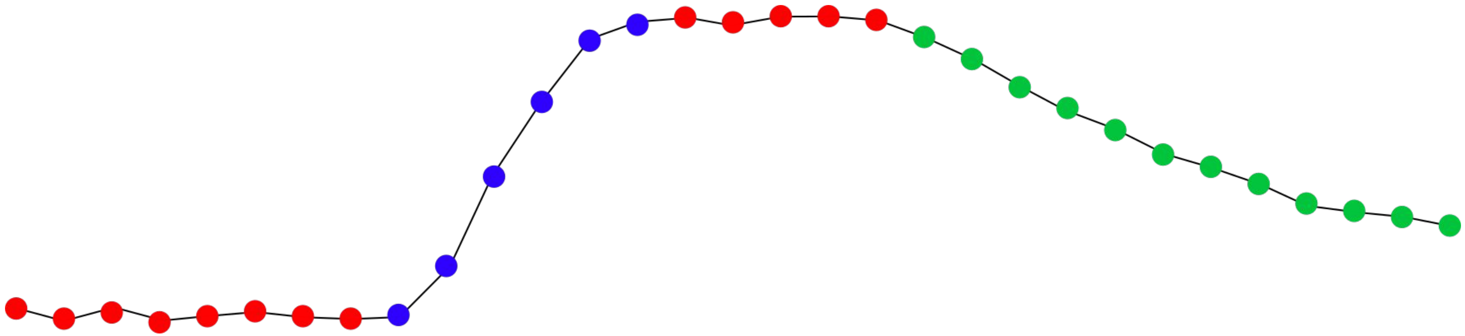
Saccades

Problem statement: 3EMCP

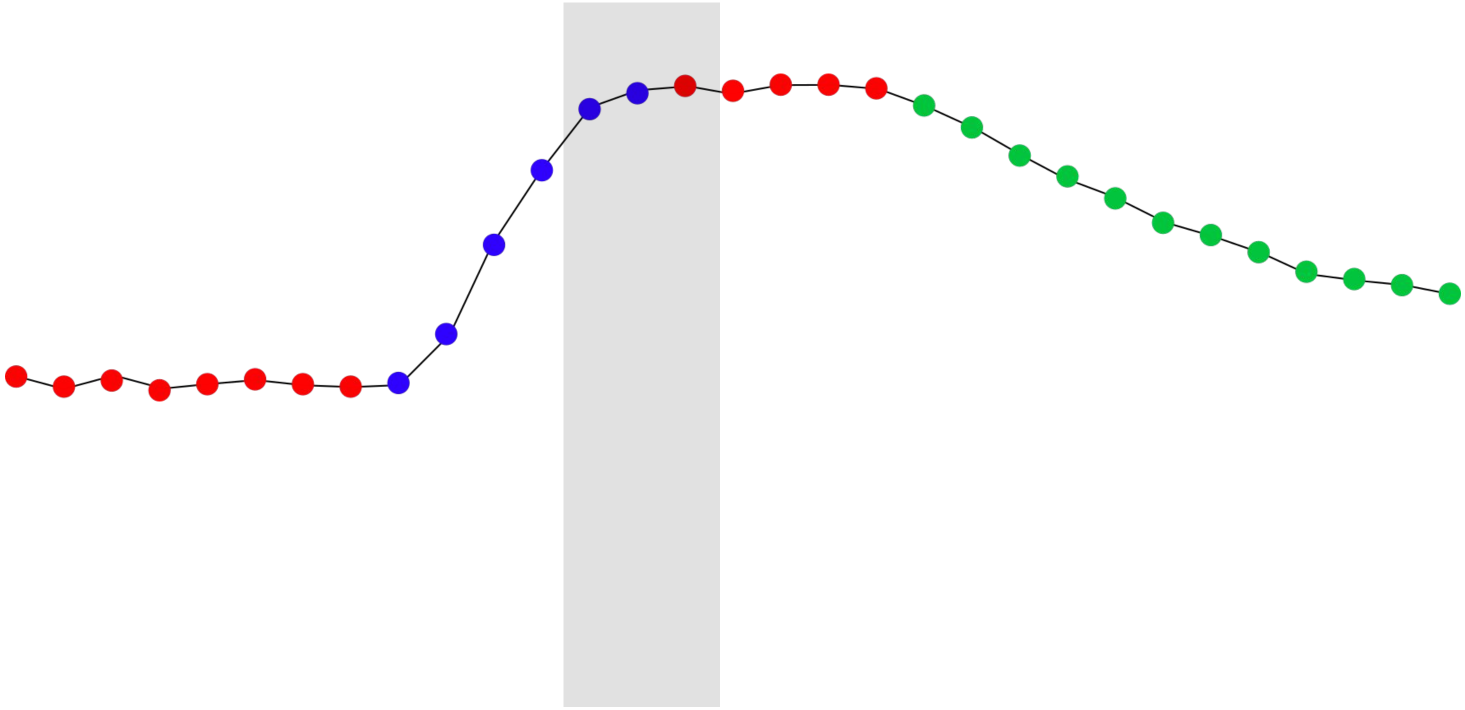


Smooth
Pursuits

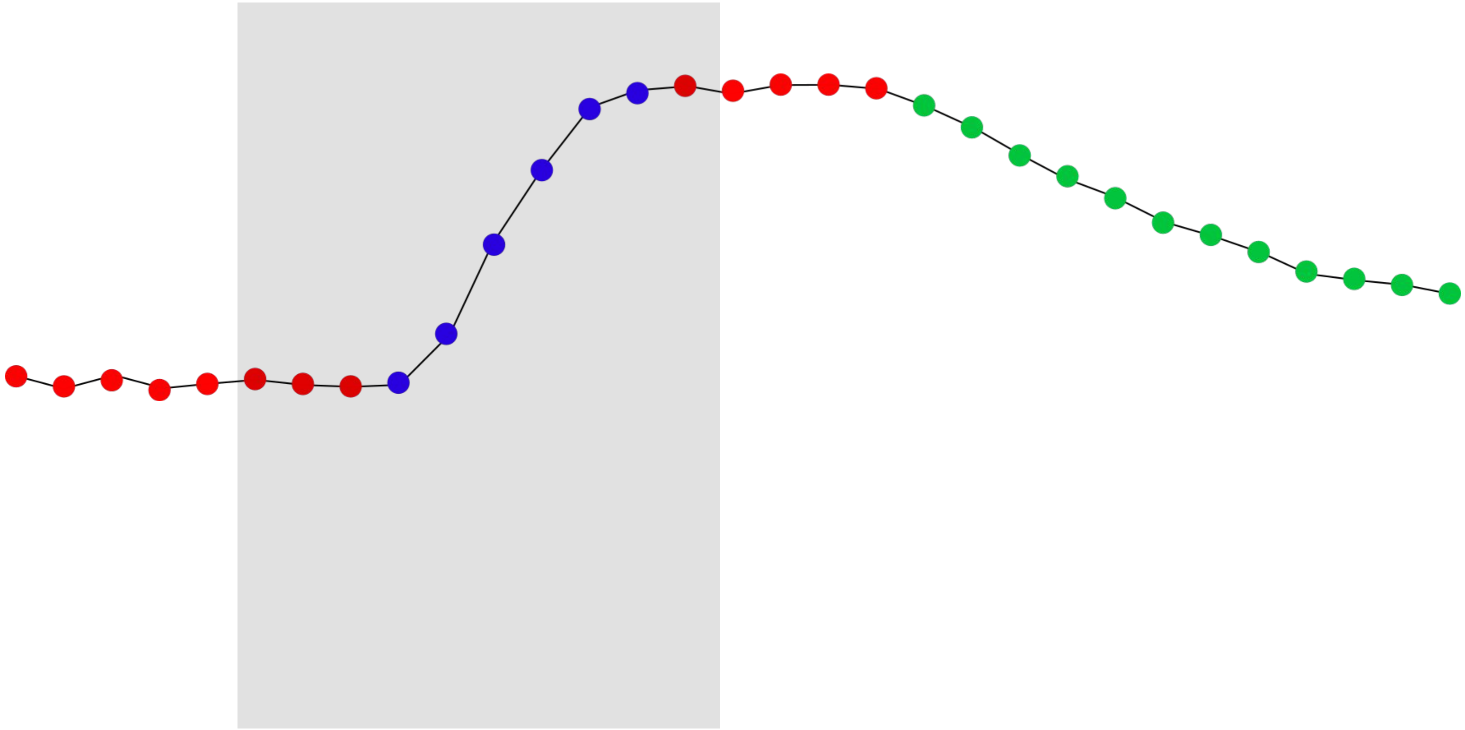
Problem statement: 3EMCP



Problem statement: 3EMCP



Problem statement: 3EMCP



Past works

Past works

- threshold-based classification

Past works

- threshold-based classification
- probabilistic methods

Past works

- threshold-based classification
- probabilistic methods
- deep learning models

End-to-end architectures

Pros

Cons

End-to-end architectures

Pros

- high accuracy

Cons

End-to-end architectures

Pros

- high accuracy
- no parameter setting

Cons

End-to-end architectures

Pros

- high accuracy
- no parameter setting
- user independent

Cons

End-to-end architectures

Pros

- high accuracy
- no parameter setting
- user independent

Cons

- computational cost

End-to-end architectures

Pros

- high accuracy
- no parameter setting
- user independent

Cons

- computational cost
- large datasets

End-to-end architectures

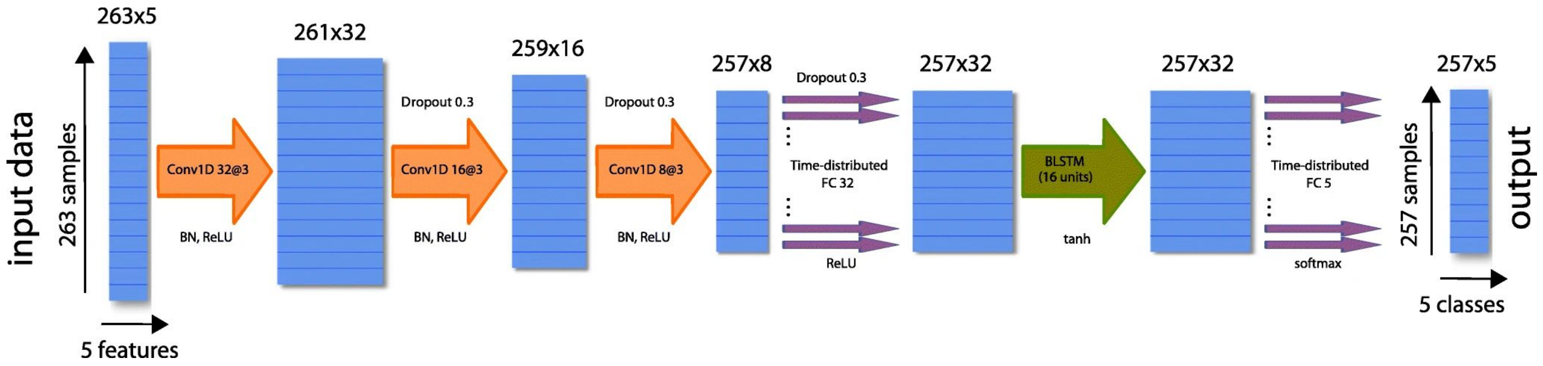
Pros

- high accuracy
- no parameter setting
- user independent

Cons

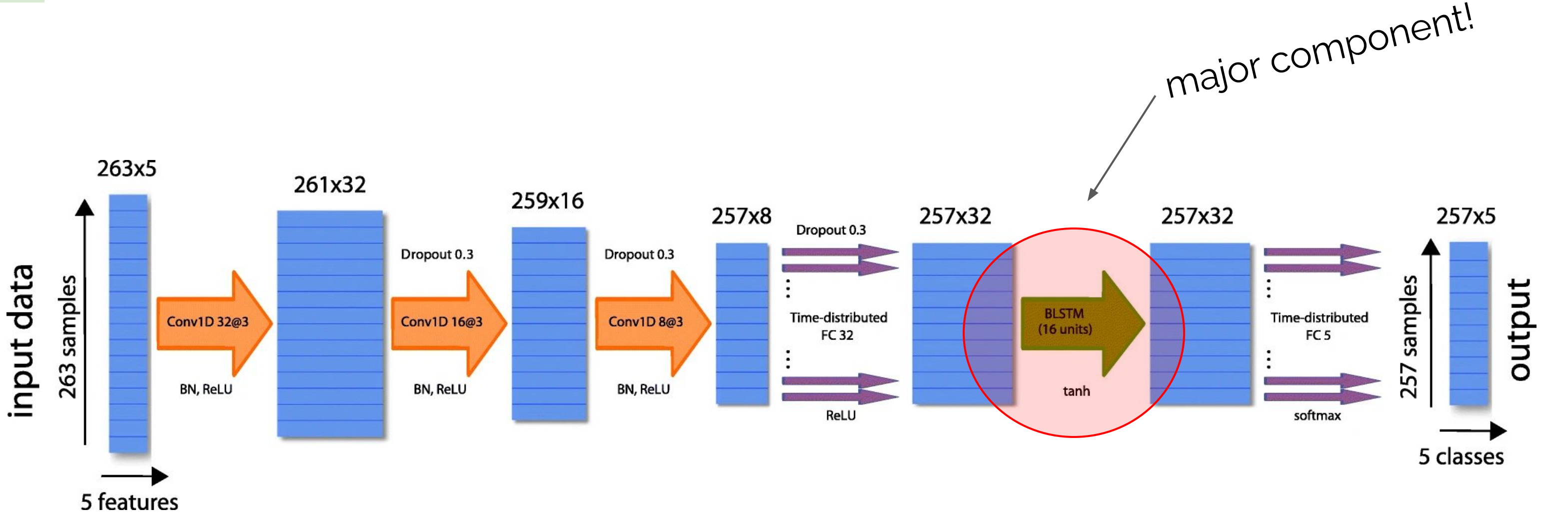
- computational cost
- large datasets
- re-training

The 1D CNN-BLSTM model



SOURCE: https://link.springer.com/article/10.3758/s13428-018-1144-2/figures/1

The 1D CNN-BLSTM model



SOURCE: https://link.springer.com/article/10.3758/s13428-018-1144-2/figures/1

Why TCNs?

Why TCNs?

- performance >= LSTMs, GRUs...

Why TCNs?

- performance >= LSTMs, GRUs...
- highly parallel

Why TCNs?

- performance >= LSTMs, GRUs...
- highly parallel
- memory footprint

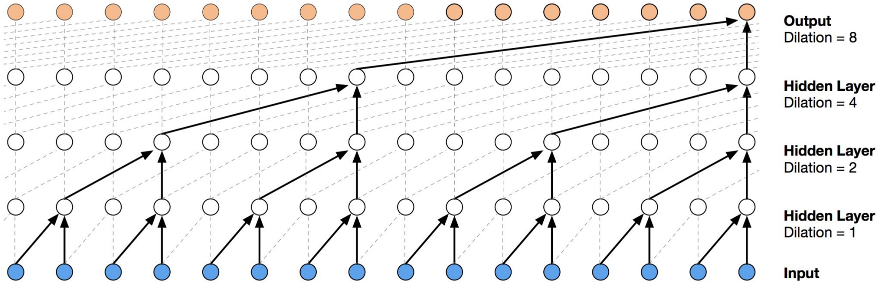
Why TCNs?

- performance >= LSTMs, GRUs...
- highly parallel
- memory footprint
- long-term dependencies

Canonical TCNs

Canonical TCNs

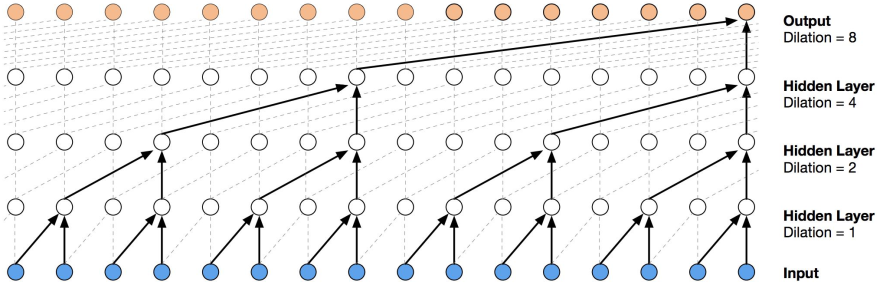
- causal convolutions



SOURCE: https://github.com/philipperemy/keras-tcn

Canonical TCNs

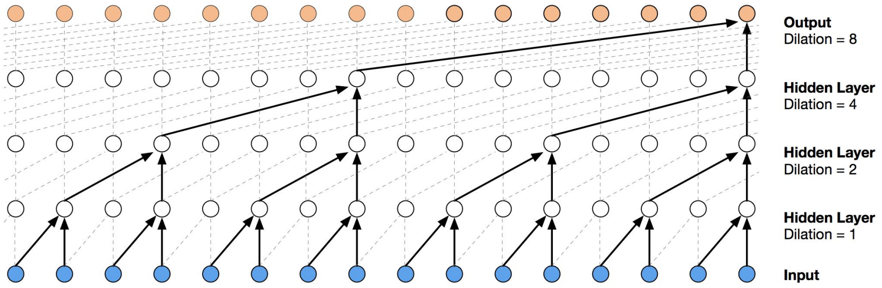
- causal convolutions
- dilations



SOURCE: https://github.com/philipperemy/keras-tcn

Canonical TCNs

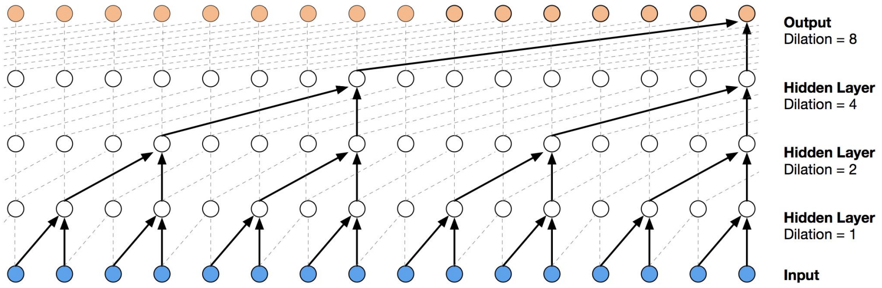
- causal convolutions
- dilations
- residual block



SOURCE: https://github.com/philipperemy/keras-tcn

Canonical TCNs

- causal convolutions
- dilations
- residual block
- variable input length

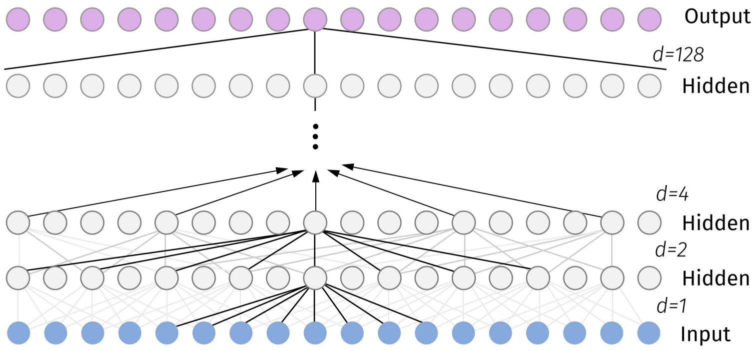


SOURCE: https://github.com/philipperemy/keras-tcn

Our TCN

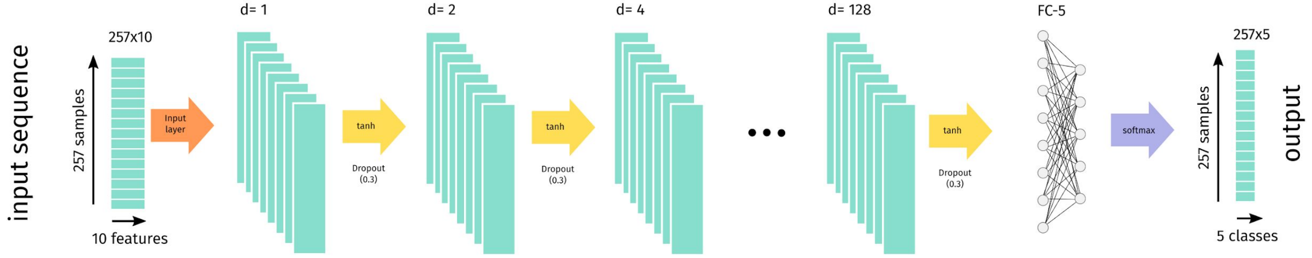
Our TCN

- non-causal convolutions



Our TCN

- non-causal convolutions
- *tanh* activation



Our TCN

- non-causal convolutions
- *tanh* activation
- specifics:

→ **https://github.com/elmadjian/3EMCP-with-TCNs**

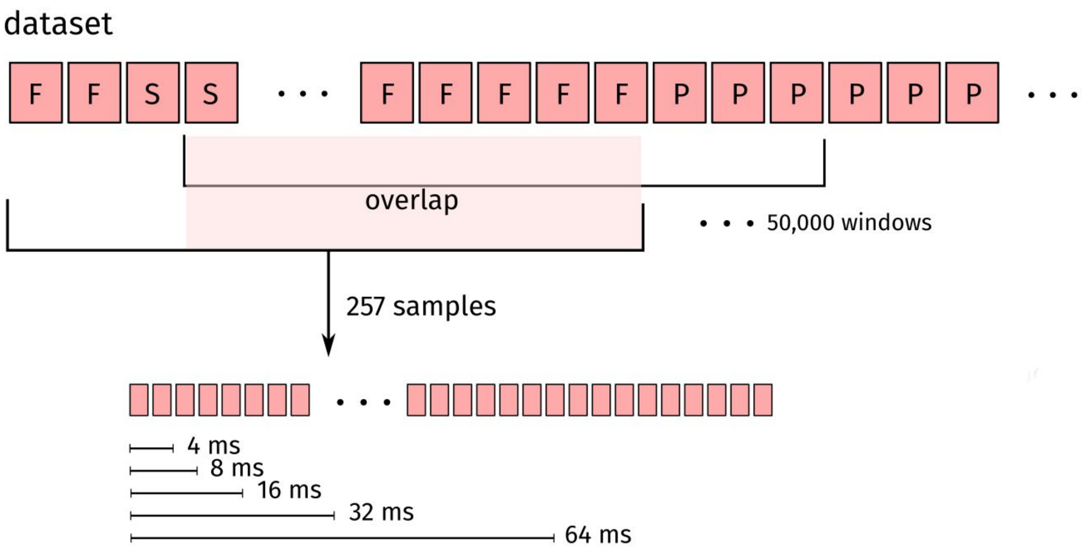
Exploratory study

Exploratory study

- feature space → *speed, direction, displacement, stddev*

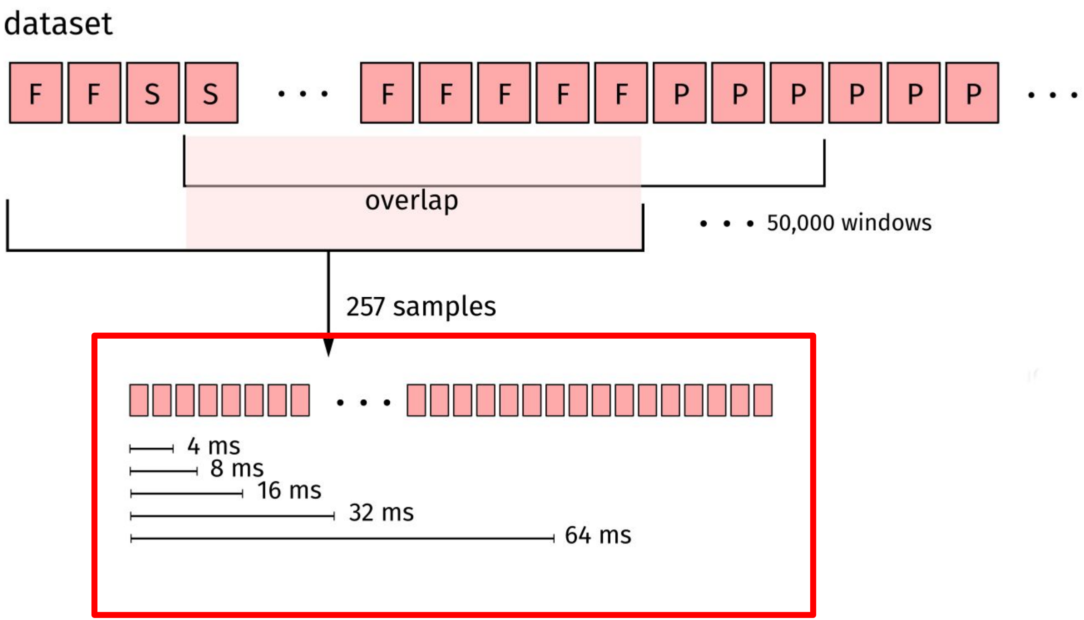
Exploratory study

- feature space
- feature scale



Exploratory study

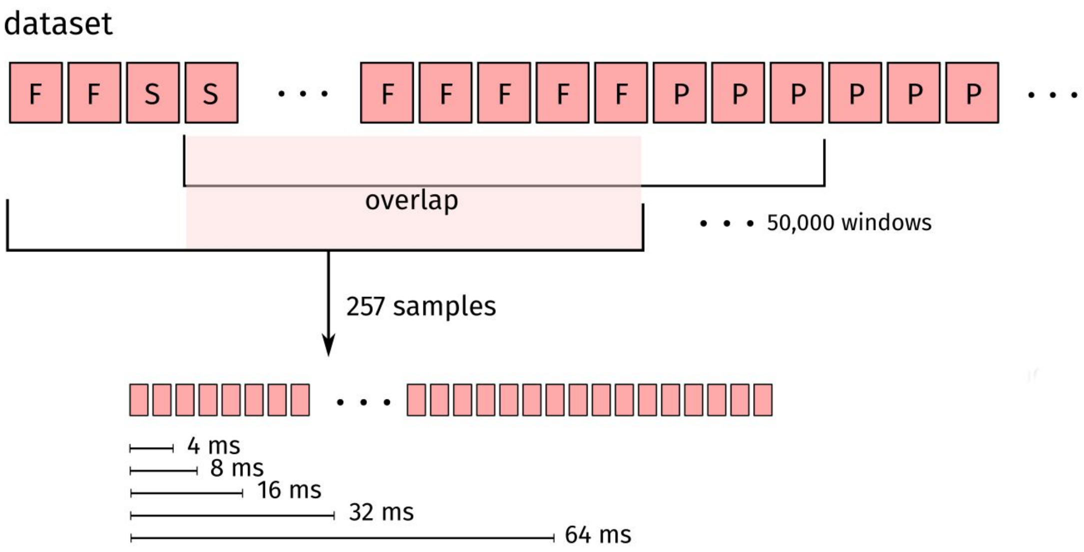
- feature space
- feature scale



+128 ms, 256 ms, 512 ms

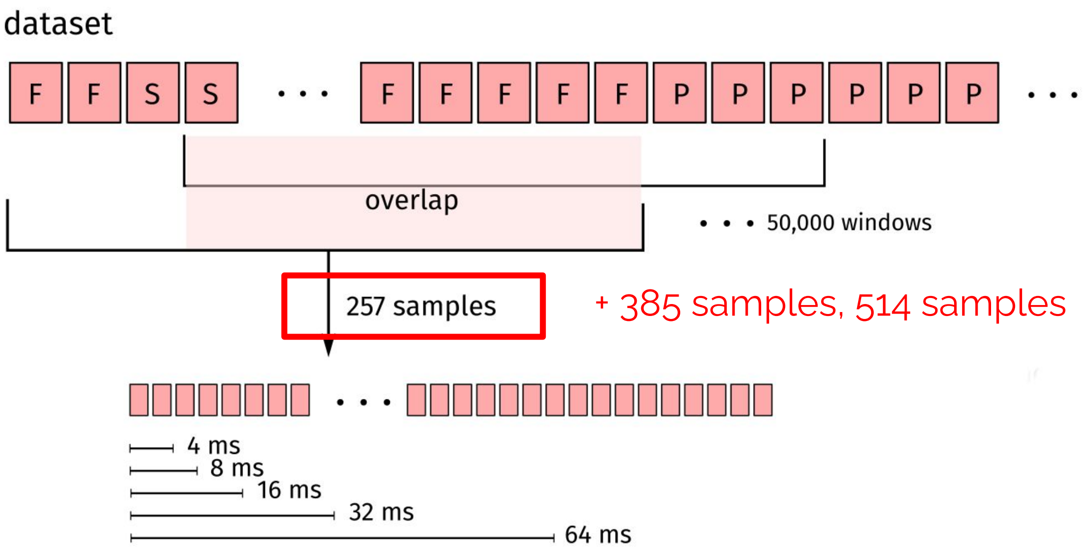
Exploratory study

- feature space
- feature scale
- window size



Exploratory study

- feature space
- feature scale
- window size



GazeCom dataset

18 clips

47 users

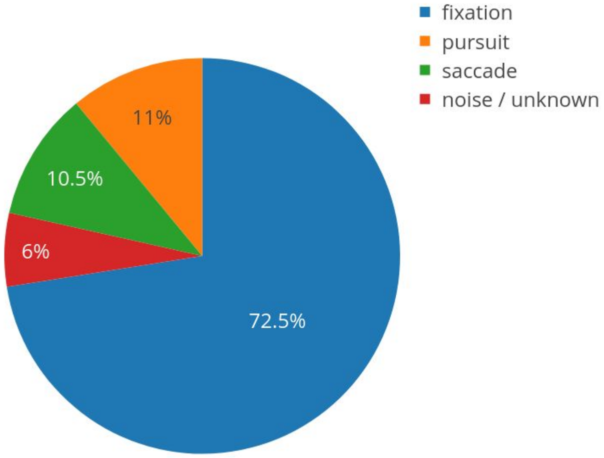
21 s

GazeCom dataset

18 clips

47 users

21 s



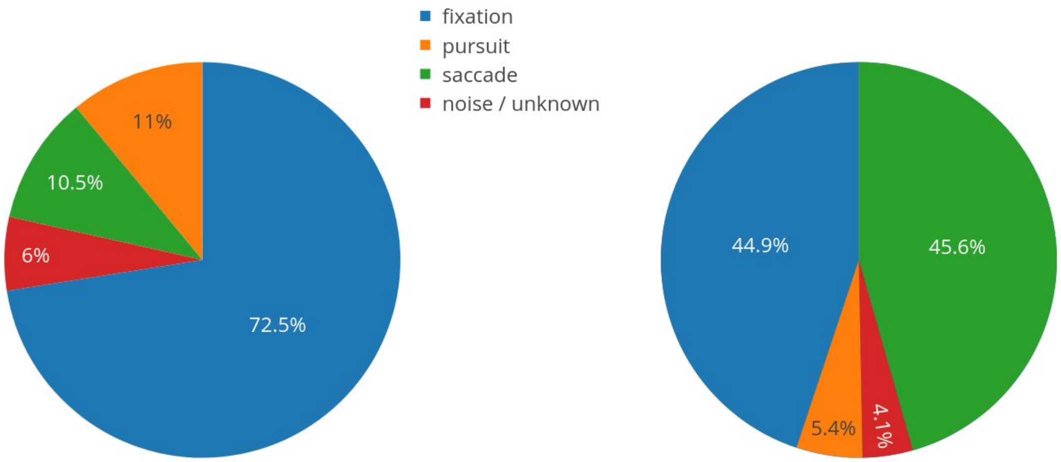
samples

GazeCom dataset

18 clips

47 users

21 s



samples

events

Training

[1] Startsev, Mikhail, Ioannis Agtzidis, and Michael Dorr. "1D CNN with BLSTM for automated classification of fixations, saccades, and smooth pursuits." *Behavior Research Methods* 51.2 (2019): 556-572.

Training

LOVO, as described by Startsev et al. [1], **except** for:

[1] Startsev, Mikhail, Ioannis Agtzidis, and Michael Dorr. "1D CNN with BLSTM for automated classification of fixations, saccades, and smooth pursuits." *Behavior Research Methods* 51.2 (2019): 556-572.

Training

LOVO, as described by Startsev et al. [1], **except** for:

- no input padding

[1] Startsev, Mikhail, Ioannis Agtzidis, and Michael Dorr. "1D CNN with BLSTM for automated classification of fixations, saccades, and smooth pursuits." *Behavior Research Methods* 51.2 (2019): 556-572.

Training

LOVO, as described by Startsev et al. [1], **except** for:

- no input padding
- ADAM optimizer (lr = 0.0001)

[1] Startsev, Mikhail, Ioannis Agtzidis, and Michael Dorr. "1D CNN with BLSTM for automated classification of fixations, saccades, and smooth pursuits." *Behavior Research Methods* 51.2 (2019): 556-572.

Training

LOVO, as described by Startsev et al. [1], **except** for:

- no input padding
- ADAM optimizer (lr = 0.0001)
- 20 epochs (large receptive field)

[1] Startsev, Mikhail, Ioannis Agtzidis, and Michael Dorr. "1D CNN with BLSTM for automated classification of fixations, saccades, and smooth pursuits." *Behavior Research Methods* 51.2 (2019): 556-572.

Training

LOVO, as described by Startsev et al. [1], **except** for:

- no input padding
- ADAM optimizer (lr = 0.0001)
- 20 epochs (large receptive field)
- batch size = 128

[1] Startsev, Mikhail, Ioannis Agtzidis, and Michael Dorr. "1D CNN with BLSTM for automated classification of fixations, saccades, and smooth pursuits." *Behavior Research Methods* 51.2 (2019): 556-572.

Training

LOVO, as described by Startsev et al. [1], **except** for:

- no input padding
- ADAM optimizer (lr = 0.0001)
- 20 epochs (large receptive field)
- batch size = 128
- tensorflow 2

Metrics and evaluation

Metrics and evaluation

- sp_tool

Metrics and evaluation

- sp_tool
- F1-score (samples and events)

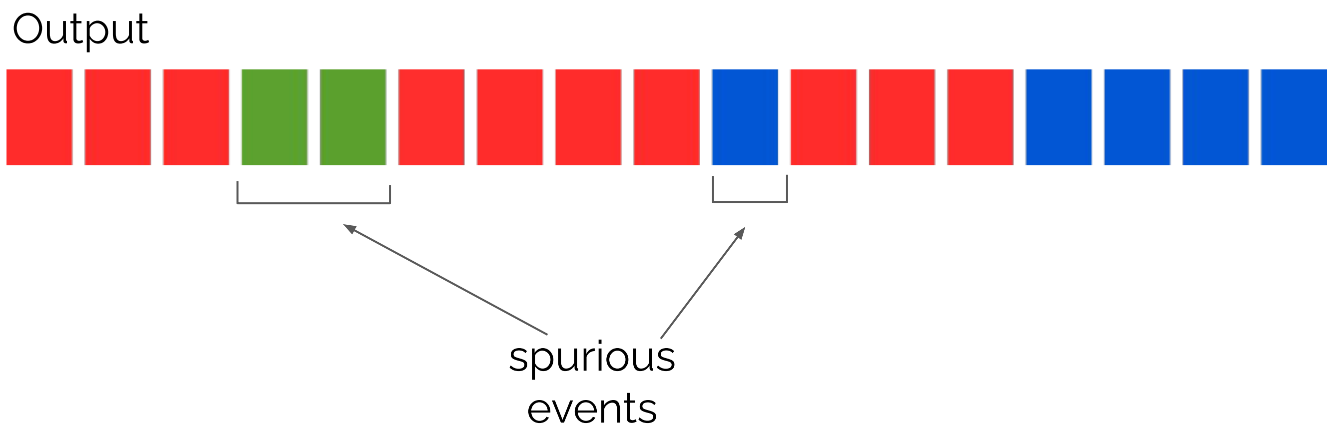
Metrics and evaluation

- sp_tool
- F1-score (samples and events)
- Events according to Hooge et al. [2]

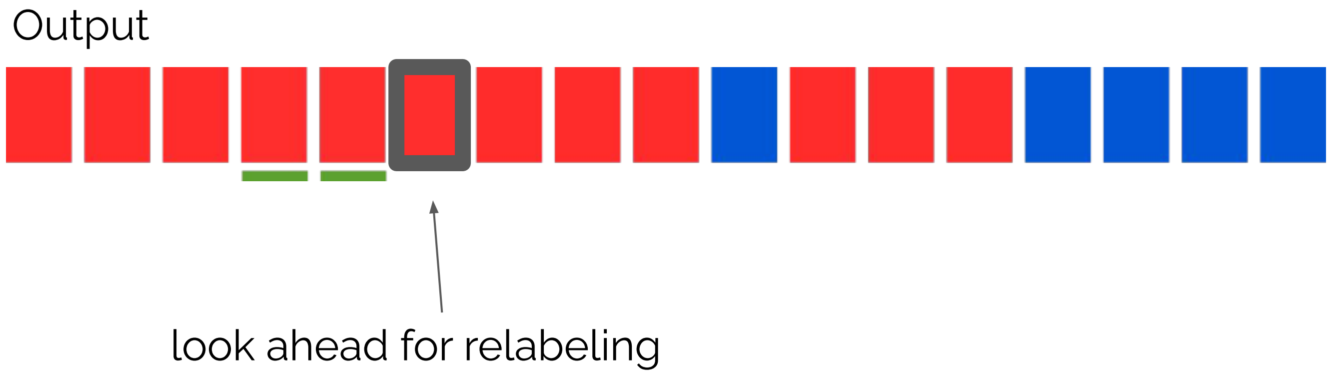
[2] Hooge, Ignace TC, et al. "Is human classification by experienced untrained observers a gold standard in fixation detection?." *Behavior Research Methods* 50.5 (2018): 1864-1881.

Knowledge-based filtering

Knowledge-based filtering



Knowledge-based filtering



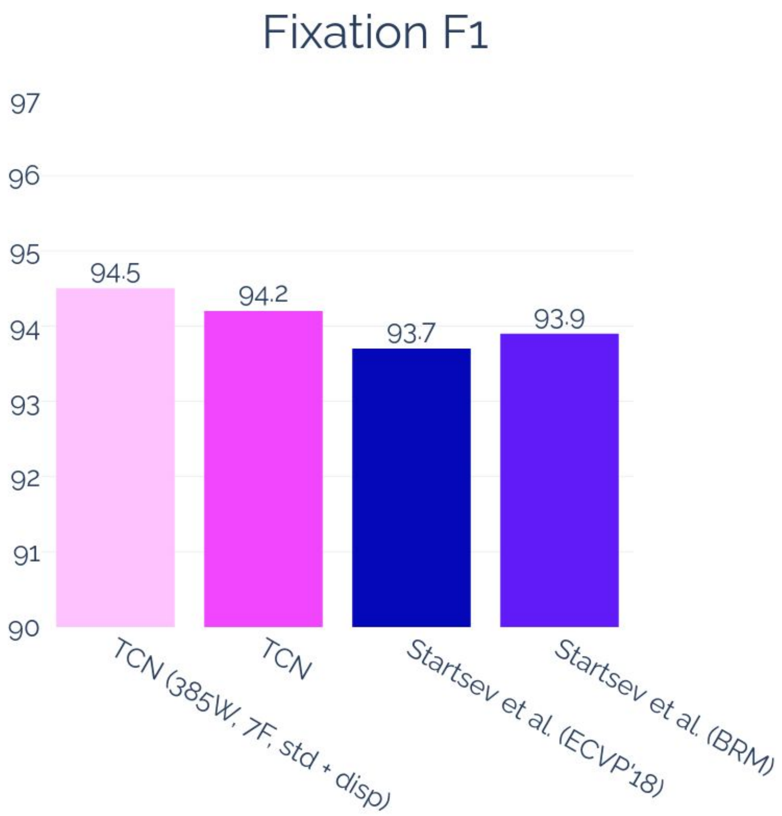
Knowledge-based filtering

Output filtered

Results

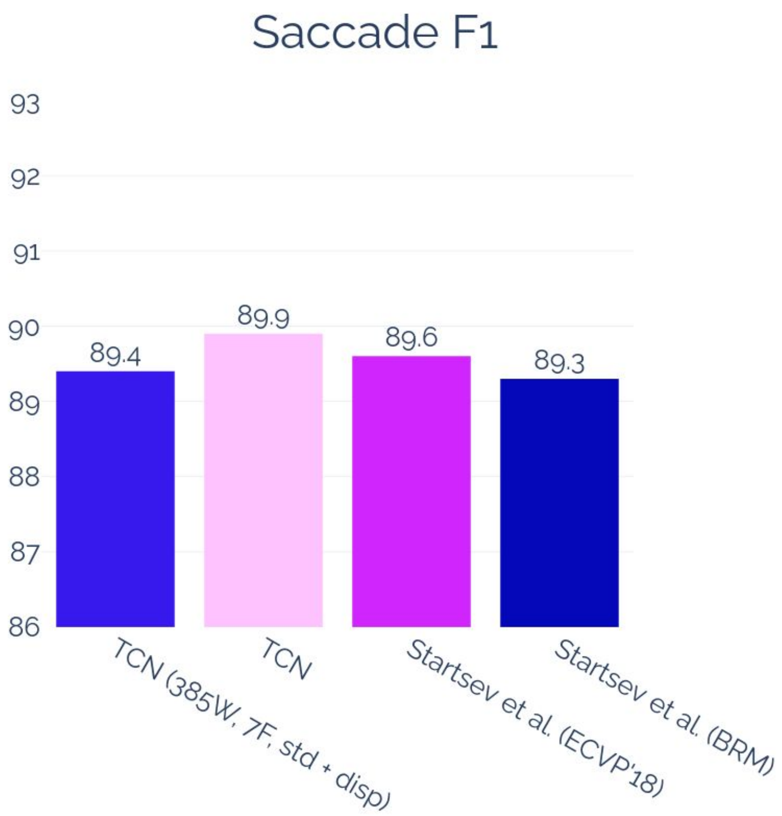
Results

- sample level



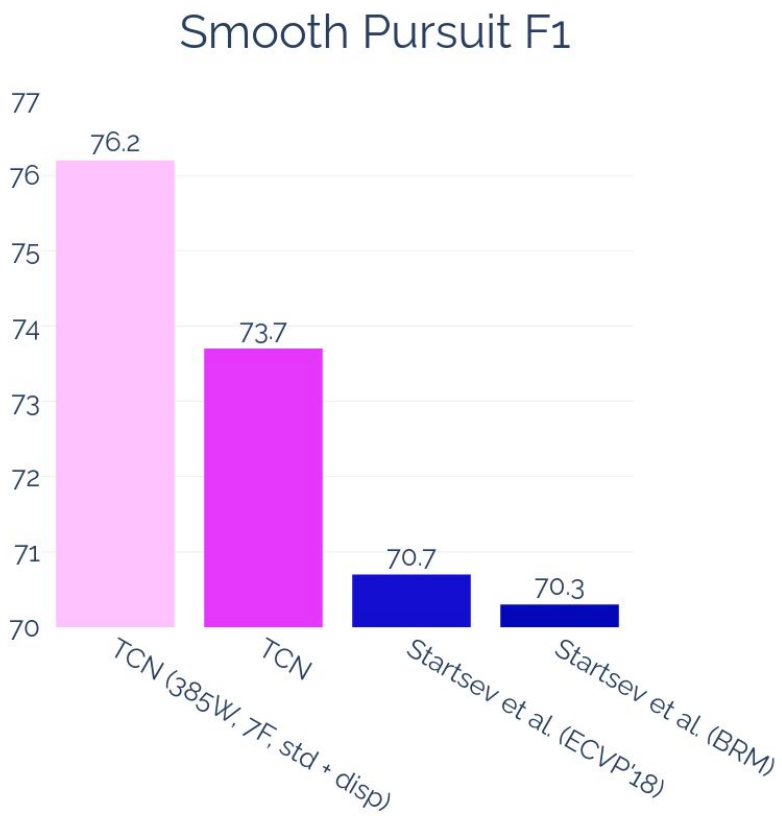
Results

- sample level



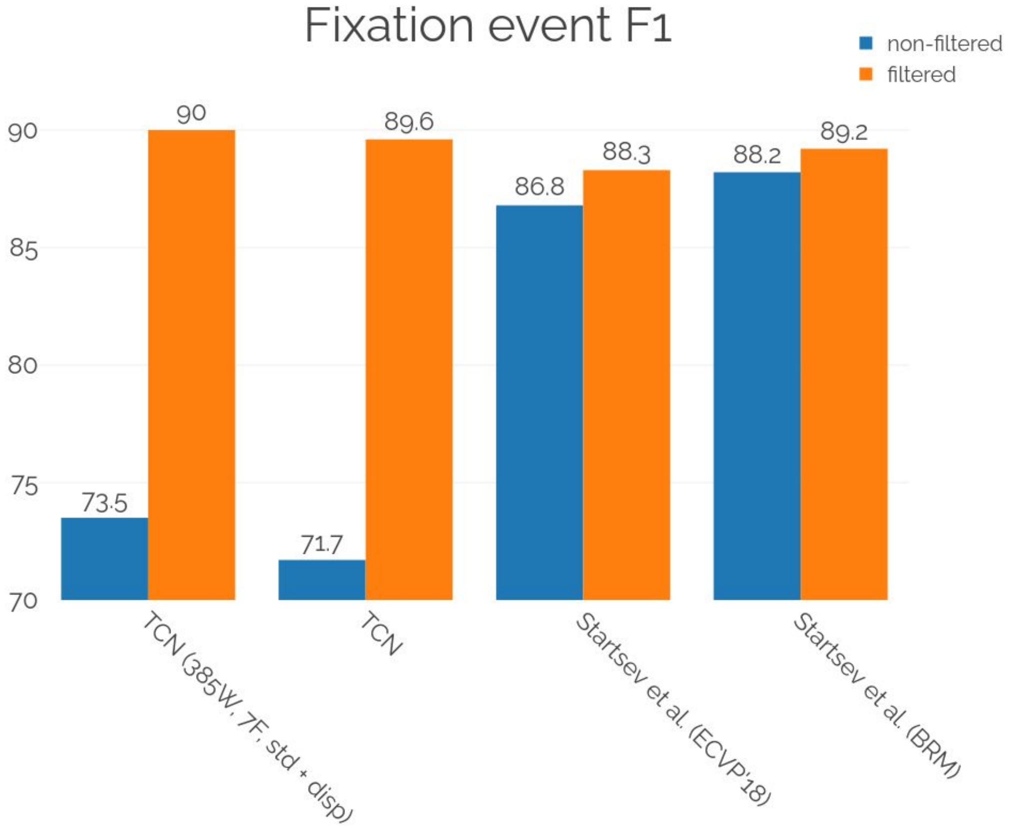
Results

- sample level



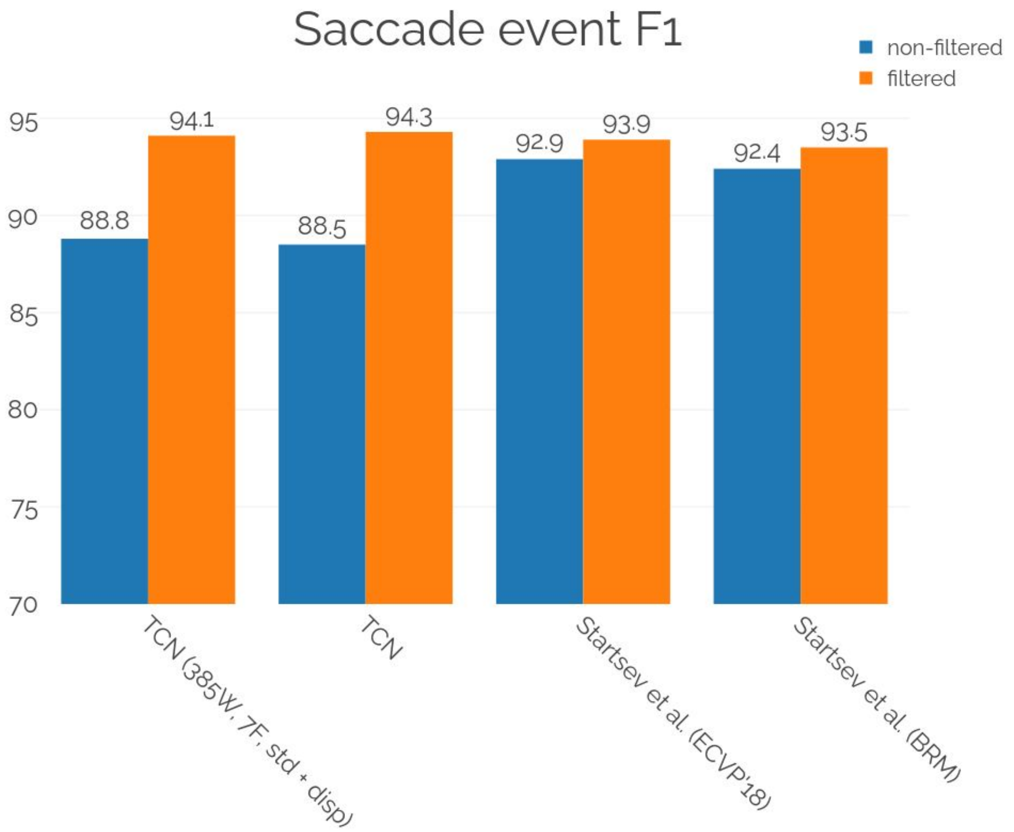
Results

- sample level
- event level



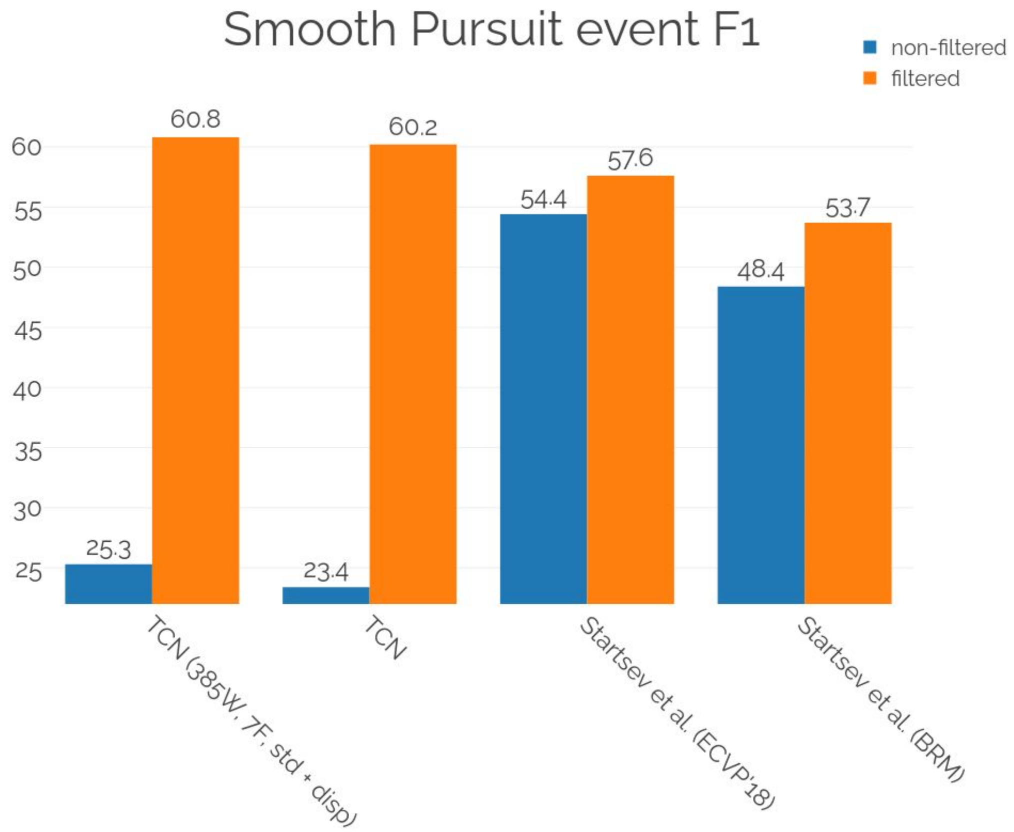
Results

- sample level
- event level



Results

- sample level
- event level



Results

- sample level
- event level
- exploratory

Results

feature space: +0.4% in SP F1

- sample level
- event level
- exploratory

Results

- sample level
- event level
- exploratory

feature space: +0.4% in SP F1

feature scale: +0.9% in SP F1 (256 ms)

Results

- sample level
- event level
- exploratory

feature space: +0.4% in SP F1

feature scale: +0.9% in SP F1 (256 ms)

window size: +1.4% in SP F1 (385 samples)

Takeaways

Takeaways

- TCNs = improvement on the state of the art

Takeaways

- TCNs = improvement on the state of the art
- feature expansion helps, but slightly

Takeaways

- TCNs = improvement on the state of the art
- feature expansion helps, but slightly
- architectural advantages

Takeaways

- TCNs = improvement on the state of the art
- feature expansion helps, but slightly
- architectural advantages
- **future work:** online classification

Takeaways

- TCNs = improvement on the state of the art
- feature expansion helps, but slightly
- architectural advantages
- **future work:** online classification

→ https://github.com/elmadjian/3EMCP-with-TCNs

Q&A