Action Spotting
Soccer Net Dataset

Authors:
Siddhesh Shinde
MS CSE
Harshit Monish
MS AI

University at Buffalo
May 11th, 2022

University at Buffalo
Department of Computer Science and Engineering
School of Engineering and Applied Sciences

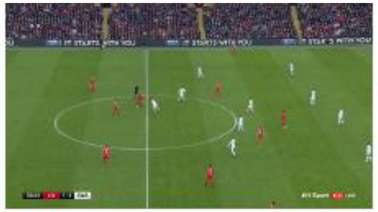
INTRODUCTION

Problem statement

- To **find all the actions/events** occurring in the Soccer Game.
- Addresses the problem of **retrieving moments with a specific semantic meaning** in long untrimmed videos.
- Contains **17 classes**
- Each action is associated with **single timestamp**.

ACTIONS


Foul


Indirect free-kick


Clearance


Shot on target


Shot off target


Corner


Substitution


Kick-off


Yellow card


Offside


Direct free-kick


Goal


Ball out of play


Throw-in


Penalty


Yellow then red card


Red card

CHALLENGES

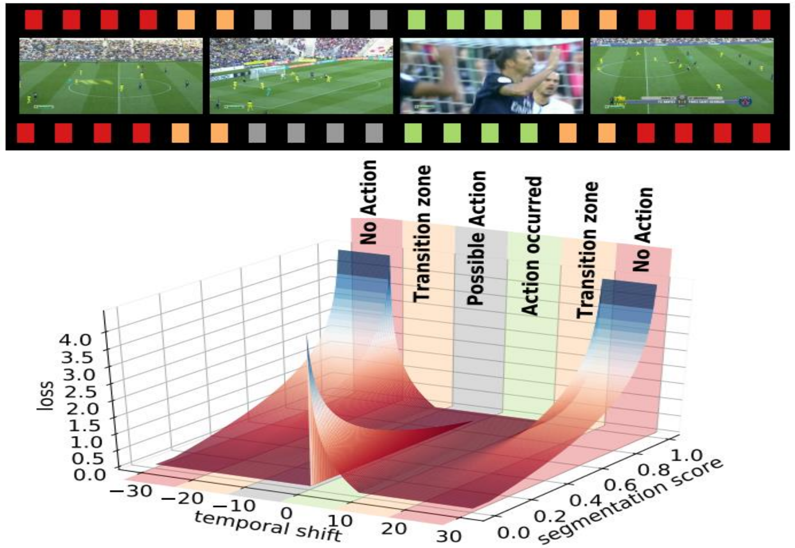
1. Sparse annotations
2. Imbalance classes
3. Classifying closely related events
4. Multiple events can resemble the same event.
5. Tight mAP (evaluation) of 5 seconds

MOTIVATION

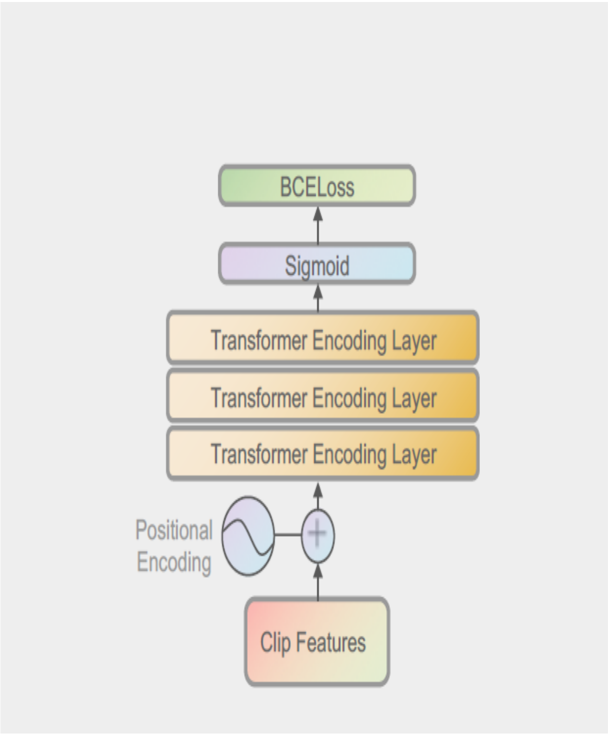
- Sports industry is a lucrative sector
- Automated analysis can help in the localization of the salient actions of a game.
- Action Spotting helps in subsequent analysis
  - replay generation
  - statistical analysis of the game/teams/players
  - Highlights generation

PREVIOUS WORK

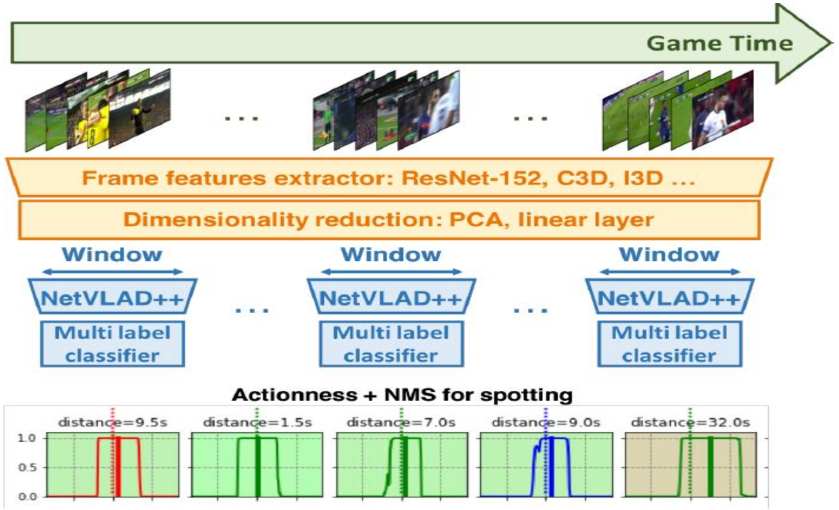
CALF



Baidu



- CALF (Context Aware Loss Function)
- NetVLAD
- NetVLAD++
  - Temporal pooling
- Baidu

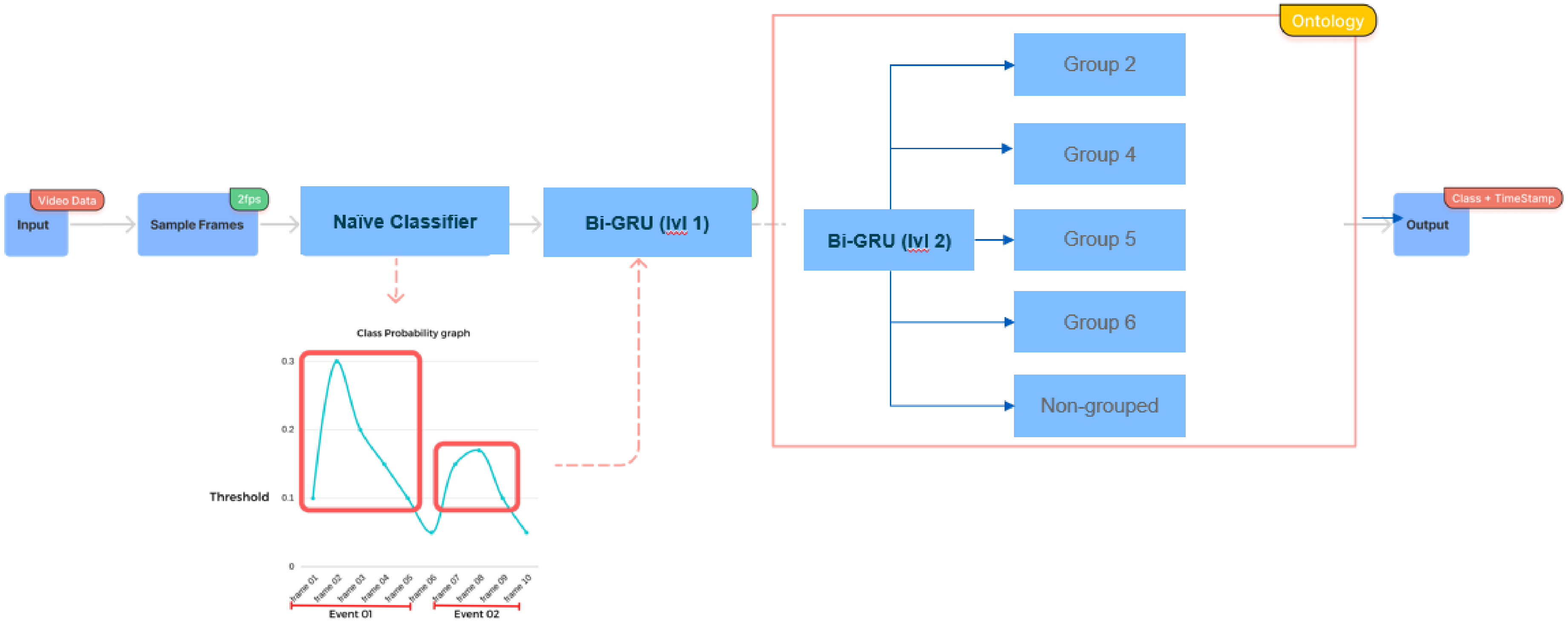


NetVLAD++

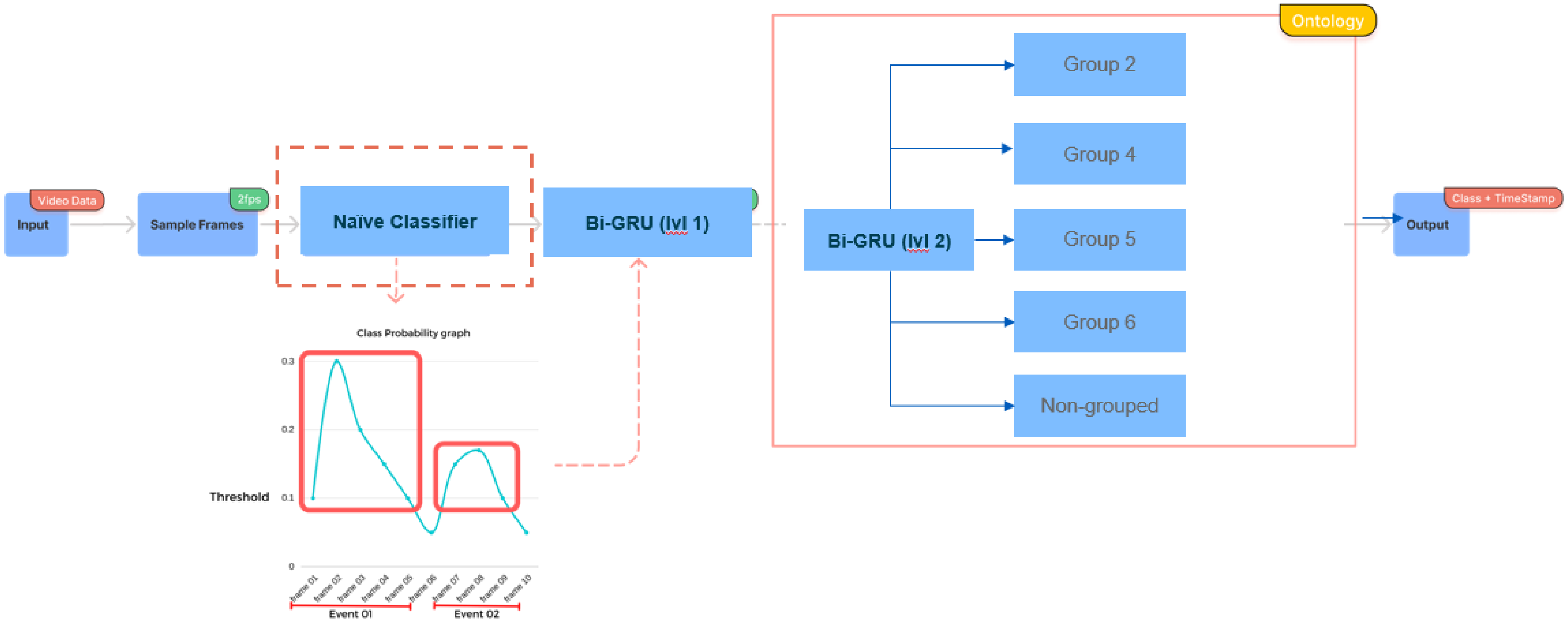
Source: https://github.com/SoccerNet/sn-spotting (05/11/22)

METHODOLOGY

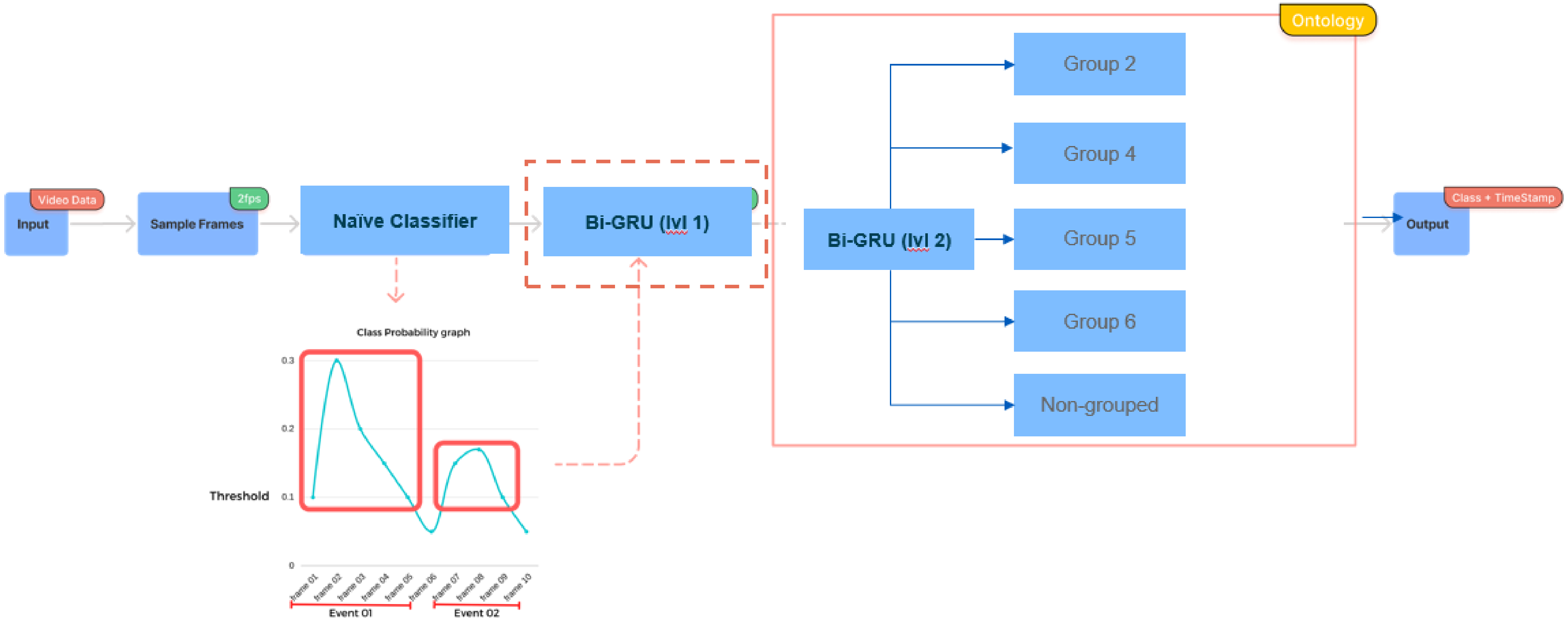
System Architecture



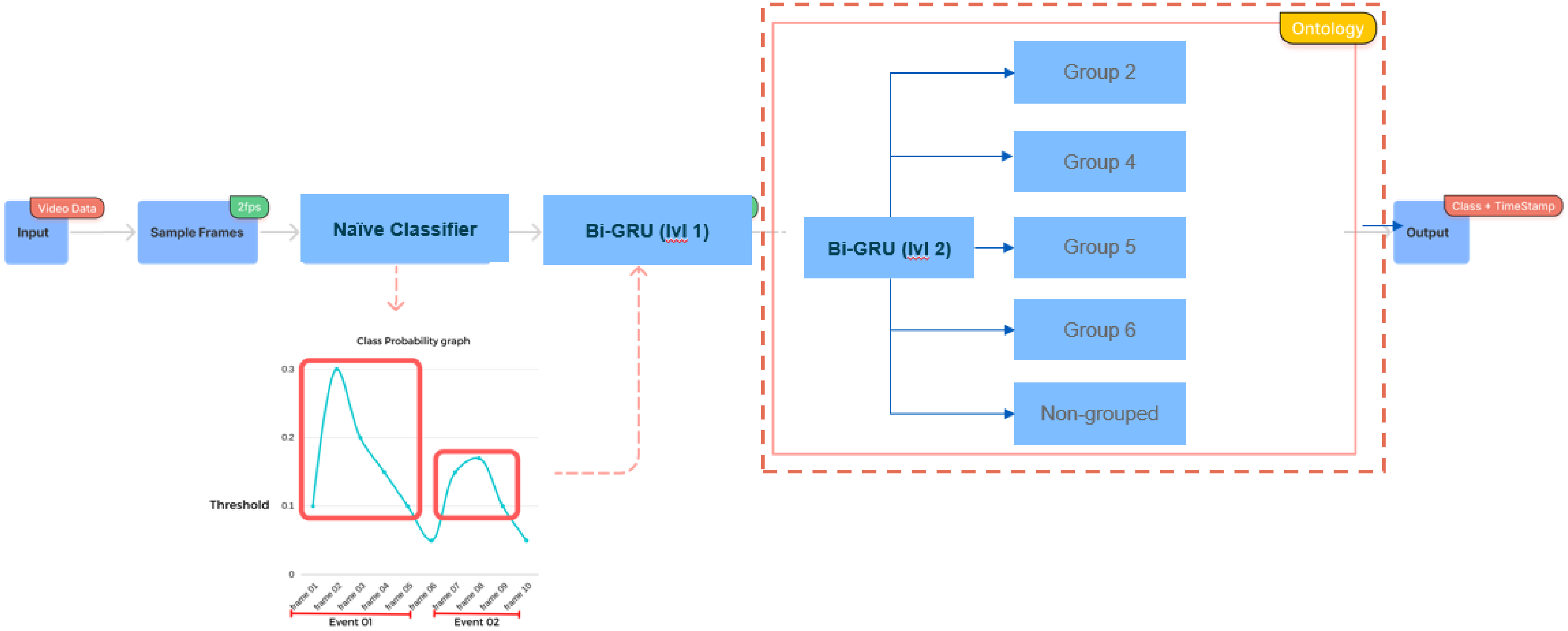
System Architecture



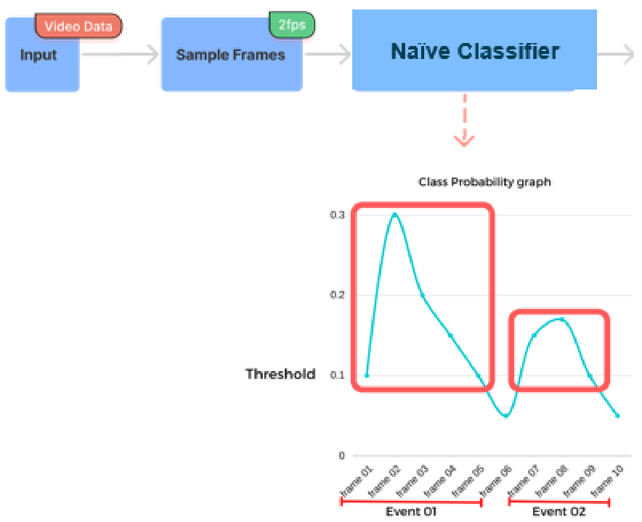
System Architecture



System Architecture



Naïve Classifier



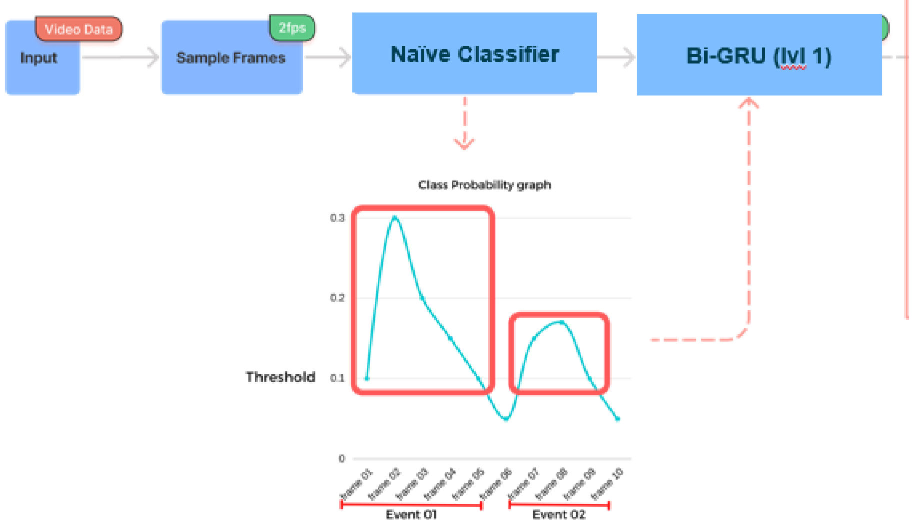
2 layer FNN
(Binary Classifier)

University at Buffalo The State University of New York

Bi-GRU (lvl 1)

```
# Event name to label index for SoccerNet-V3
EVENT_DICTIONARY_V2 = {
        "Penalty":0,"Kick-off":1,"Goal":2,"Substitution":3,
        "Offside":4,"Shots on target":5,"Shots off target":6,
        "Clearance":7,"Ball out of play":8,"Throw-in":9,"Foul":10,
    "Indirect free-kick":11,"Direct free-kick":12,"Corner":13,
    "Yellow card":14,"Red card":15,"Yellow->red card":16
}
```

```
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 bidirectional (Bidirectiona  (None, 15, 32)           1653120
 l)

 bidirectional_1 (Bidirectio  (None, 32)               12672
 nal)

 flatten (Flatten)           (None, 32)                0

 dense (Dense)               (None, 64)                2112

 dropout (Dropout)           (None, 64)                0

 dense_1 (Dense)             (None, 64)                4160

 dropout_1 (Dropout)         (None, 64)                0

 dense_2 (Dense)             (None, 17)                1105

=================================================================
Total params: 1,673,169
Trainable params: 1,673,169
Non-trainable params: 0
_________________________________________________________________
```

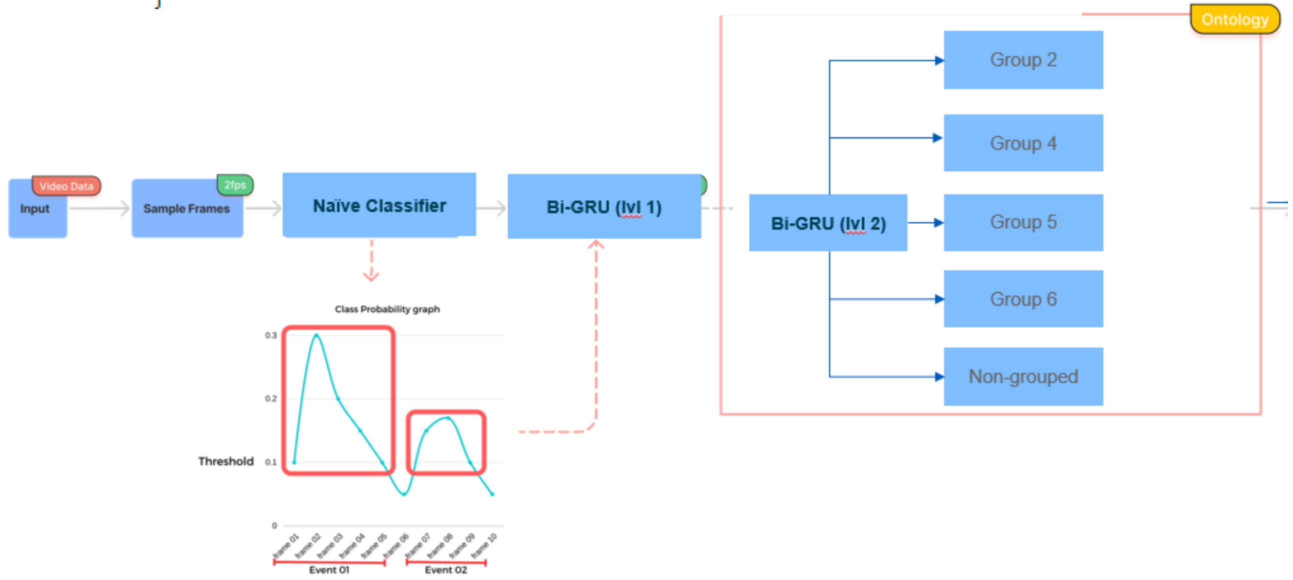


Event size is 15 frames (secs)

Bi-GRU (lvl 2)

```
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 bidirectional_2 (Bidirectio  (None, 11, 128)          6686208
 nal)

 dropout_2 (Dropout)         (None, 11, 128)           0

 bidirectional_3 (Bidirectio  (None, 128)              198144
 nal)

 dropout_3 (Dropout)         (None, 128)               0

 flatten_1 (Flatten)         (None, 128)               0

 dense_1 (Dense)             (None, 8)                 1032

=================================================================
Total params: 6,885,384
Trainable params: 6,885,384
Non-trainable params: 0
_________________________________________________________________
```

```
MOD_EVENT_DICTIONARY_V2 = {
        "Penalty":0,
        "Kick-off":1,
        "Goal":2,"Shots on target":2,"Shots off target":2,"Corner":2,
        "Substitution":3,
        "Yellow card":4,"Red card":4,"Yellow->red card":4,"Offside":4,
        "Clearance":5,"Ball out of play":5,"Throw-in":5,
        "Foul":6,"Indirect free-kick":6,"Direct free-kick":6,
}
```



Pre-Event and Post-Event size is 11 frames (secs)

EVALUATION

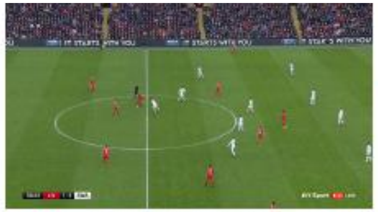
Dataset

- 764 hours of 500 games. (720p and 224p resolutions)
  - Training: 300, Validation: 100, Testing: 100
  - Challenge: 50
- Total 110,458 annotated actions on average 221 actions per game. 1 action every 25 seconds.
- Each label is marked as visible or non-visible.
- Provided BAIDU features are at 1fps.


```
"annotations": [
    {
        "gameTime": "1 - 00:00",
        "label": "Kick-off",
        "position": "0",
        "team": "home",
        "visibility": "visible"
    },
    {
        "gameTime": "1 - 01:16",
        "label": "Ball out of play",
        "position": "76376",
        "team": "not applicable",
        "visibility": "visible"
    },
    {
        "gameTime": "1 - 01:26",
        "label": "Throw-in",
        "position": "86245",
        "team": "away",
        "visibility": "not shown"
    },
    ...
]
```

Dataset


Foul


Indirect free-kick


Clearance


Shot on target


Shot off target


Corner


Substitution


Kick-off


Yellow card


Offside


Direct free-kick


Goal


Ball out of play


Throw-in


Penalty


Yellow then red card


Red card

Evaluation Metric

- The performance is assessed by the Average-mAP metric
  - A predicted action spot is positive if it falls within the given tolerance x of a ground-truth timestamp from the same class.
- The Average Precision (AP) based on PR curves is computed and then averaged over the classes(mAP), after which the Average-mAP is the AUC of the mAP at different tolerances x.

Tolerance for tight bounds: 5
Tolerance for loose bounds: 60

RESULTS

Results on tight bounds (5 seconds)

Model	tight avg. mAP (test)	tight avg. mAP (challenge)
Baidu	47.05	49.56
NetVLAD++ with Baidu features	NA	43.99
Ours	**35.17**	**36.71**
AImageLab – RMS	28.83	27.69
CALF-calibration	NA	15.83
CALF	NA	15.33
NetVLAD++	11.51	9.91
NetVLAD	4.20	4.31

Source: https://www.soccer-net.org/tasks/action-spotting (05/11/22)

Results on loose bounds (60 seconds)

Model	loose avg. mAP (test)	loose avg. mAP (challenge)
Baidu	73.77	78.84
NetVLAD++ with Baidu features	NA	74.63
Ours	**49.86**	**51.35**
AImageLab – RMS	28.83	27.69
NetVLAD++	53.4	52.54
CALF-calibration	46.80	46.39
CALF	NA	42.22
NetVLAD	31.37	30.74

Source: https://www.soccer-net.org/tasks/action-spotting (05/11/22)

Average mAP on Test Dataset

Class	mAP	Class	
Penalty	0.88	Throw-in	0.67
Kick-off	0.55	Foul	0.65
Goal	0.56	Indirect free-kick	0.45
Substitution	0.59	Direct free-kick	0.53
Offside	0.52	Corner	0.69
Shots on target	0.14	Yellow card	0.57
Shots off target	0.33	Red card	0.09
Clearance	0.53	Yellow->red card	0.5
Ball out of play	0.60		

Source: https://www.soccer-net.org/tasks/action-spotting (05/11/22)

Average mAP on Challenge Dataset

```
Task: spotting / Split: challenge / Average-mAP: 36.71 / 51.36
output
{'result': [{'challenge_split': {'Average-mAP (loose)': 51.358695424469225, 'Shown only (loose)':
55.294152573913905, 'Unshown only (loose)': 35.33511889274671, 'Average-mAP (tight)': 36.711555909310064,
'Shown only (tight)': 39.32525735323055, 'Unshown only (tight)': 21.26342508335424}}],
'submission_result': {'challenge_split': {'Average-mAP (loose)': 51.358695424469225, 'Shown only
(loose)': 55.294152573913905, 'Unshown only (loose)': 35.33511889274671, 'Average-mAP (tight)':
36.711555909310064, 'Shown only (tight)': 39.32525735323055, 'Unshown only (tight)': 21.26342508335424}}}
```

Source: https://www.soccer-net.org/tasks/action-spotting (05/11/22)

CONCLUSION AND FUTURE WORK

- **An ontology-based solution** was proposed for **Action Spotting in Soccer videos**.
- The experimental results show that the **proposed scheme can incorporate pre-event and post-event frames for classification task, especially for closely related events.**
- The reason for Yellow Card and Yellow->red card event's poor accuracy may be because of the **imbalance in the dataset.**
- The reason for Shot on Target and Shot off Target event's poor accuracy may be because of the **closeness with Goal event.**
- Further improvement in performance can be obtained by **exploring different solutions for sparsity problem** and **incorporating vision transformer for action spotting.**

QUESTIONS?

THANK YOU!