A Coarse-to-Fine Boundary Localization method for Naturalistic Driving Action Recognition

Guanchen Ding[1,*] , Wenwei Han[2,*] , Chenglong Wang[3,*] , Mingpeng Cui[1] , Lin Zhou[1] , Dianbo Pan[1] ,
Jiayi Wang[1] , Junxi Zhang[1] , Zhenzhong Chen[1,2,†]

[1]School of Remote Sensing and Information Engineering, Wuhan University, China
[2]School of Computer Science, Wuhan University, China
[3]School of Resource and Environmental Sciences, Wuhan University, China

2022 AI City Challenge

Contents

1

Problem Statements

2

Methodology

Dataset

3

Experimental Results

S. No	Distracted Driver Behavior
1	Normal Forward Driving
2	Drinking
3	Phone Call(Right)
4	Phone Call(Left)
5	Eating
6	Text(Right)
7	Text(Left)
8	Hair / Makeup
9	Reaching Behind
10	Adjust Control Panel
11	Pick Up From Floor(Driver)
12	Pick Up From Floor(Passenger)
13	Talk To Passenger At The Right
14	Talk To Passenger At Backseat
15	Yawning
16	Hand On Head
17	Singing With Music
18	Shaking Or Dancing With Music

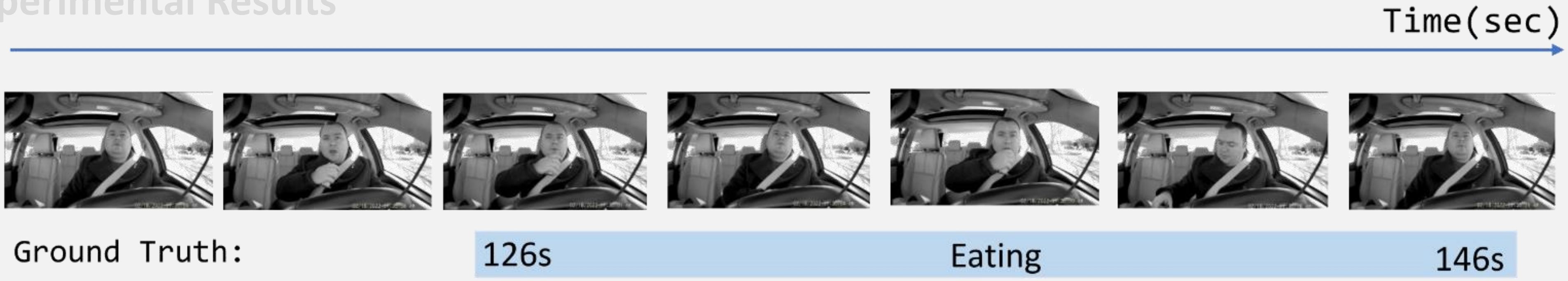
SynDD1.

Benchmark for the challenge model.

Each participant continuously performed eighteen distracted driver behavior for a small-time interval.

Problem Statements

Challenges



- Large intra-class variation

The intra-class variation may puzzle the model to divide one action segment into different parts especially for the anchor-based approaches which produce a bunch of redundant proposals.

- Multiple camera views

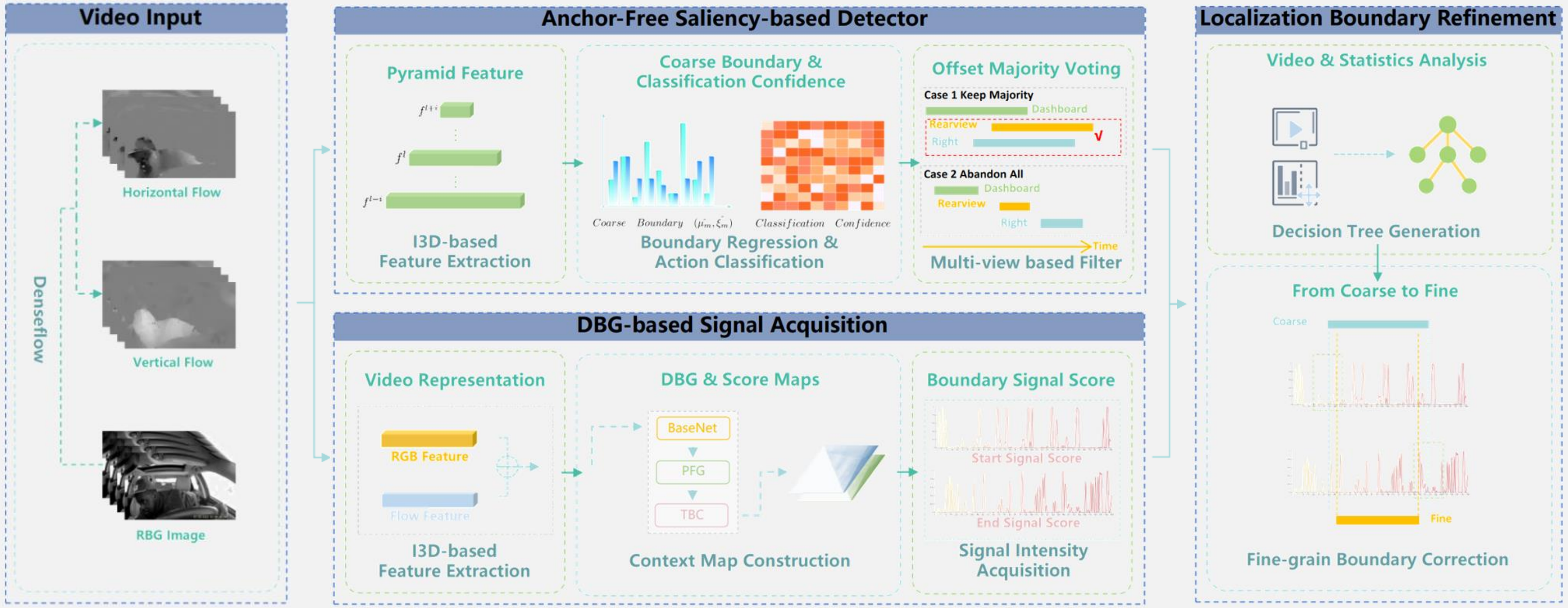
It is challenging to effectively combine the information of multiple camera views to improve the accuracy of action classification and boundary localization.

2 Methodology

3 Experimental Results

1 Background Settings

Pipeline



Our Pipeline. Given a video, our model first utilizes DenseFlow to extract optical flows, whose results, together with original RGB frames, are regarded as the inputs of the network. Then an Anchor-Free Saliency-based Detector(AFSD) is applied to obtain the classification result and coarse boundary prediction. What's more, a DBG-based Signal Acquisition Module is designed to model starting and ending signals. Finally, combining the two results and using the Localization Boundary Refinement Module as an auxiliary, the fine boundary is obtained.

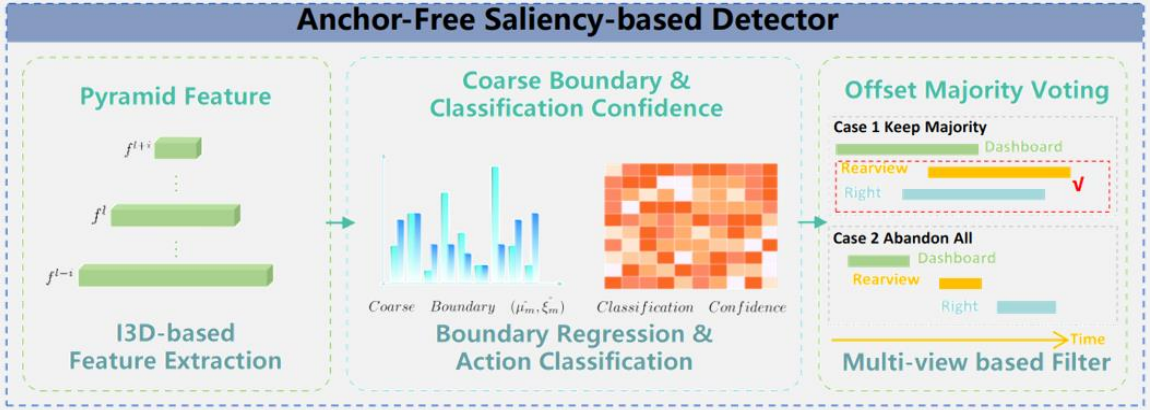
2

Methodology

3 Experimental Results

1 Background Settings

Anchor-Free Saliency-based Detector[1]

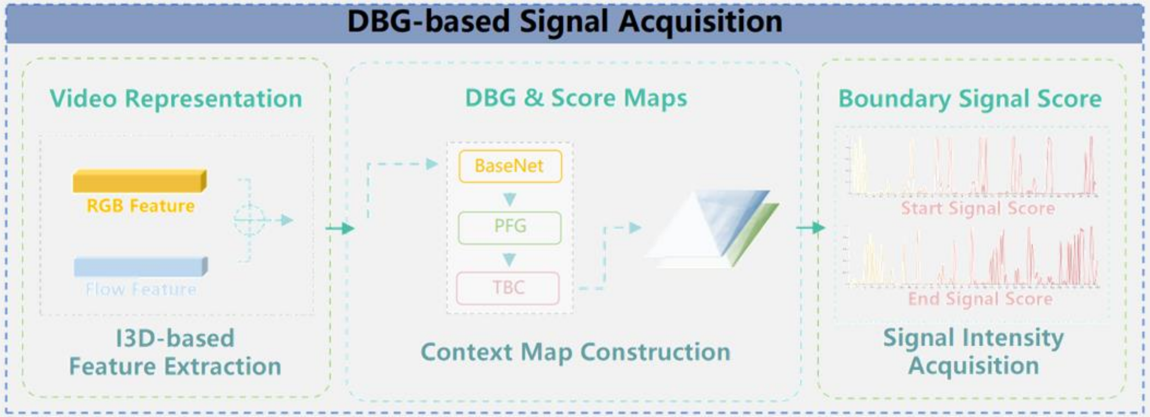


Implementation.

I3D-based Feature Extraction ~ Pyramid Feature
Boundary Regression ~ Coarse Boundary
Action Classification ~ Classification Confidence
Multi-view based Filter ~ Offset Majority Voting

[1] *Chuming Lin, Chengming Xu, Donghao Luo, Yabiao Wang, Ying Tai, Chengjie Wang, Jinlin Li, Feiyue Wang, and Yanwei Fu. Learning salient boundary feature for anchor-free temporal action localization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.*

DBG-based Signal Acquisition[2]



Motivation for DBG-based Signal Acquisition.

- aim to obtain accurate temporal boundaries.

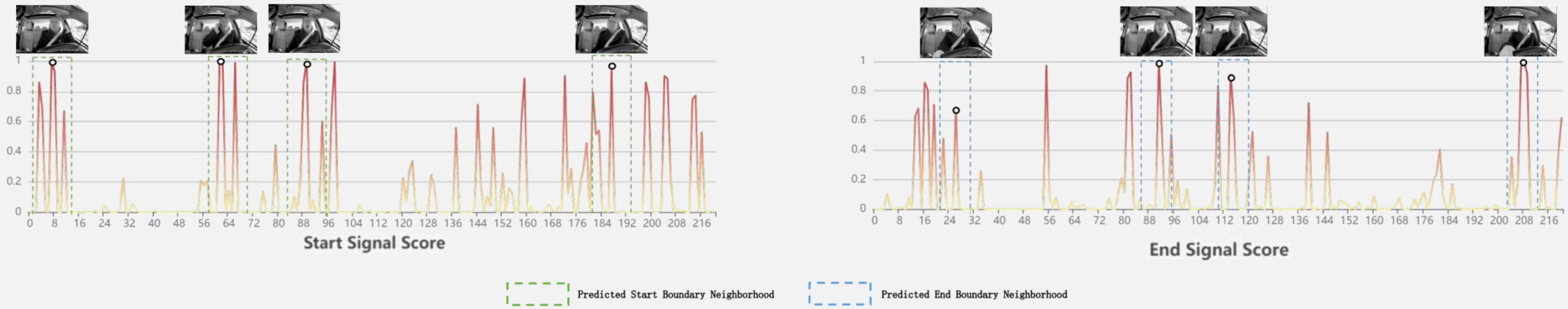
Due to the parallel optimization of the action classification task and boundary regression task, the boundary proposals we got from the previous model are trade-off products that can still be optimized.

Implementation.

I3D-based Feature Extraction ~ Video Representation
Context Map Construction ~ Score Maps
Signal Intensity Acquisition ~ Boundary Signal Score

[2] *Chuming Lin, Jian Li, Yabiao Wang, Ying Tai, Donghao Luo, Zhipeng Cui, Chengjie Wang, Jilin Li, Feiyue Huang, and Rongrong Ji. Fast learning of temporal action proposal via dense boundary generator. In proceedings of the AAAI Conference on Artificial Intelligence.*

Localization Boundary Refinement



The process of searching for strongest signal.

The real boundary lies in the neighbor of the previous predicted proposal and achieves a high score. After obtaining the coarse boundary prediction, we search the pre-defined neighbors in order to get the strongest signal that indicates the starting or ending points.

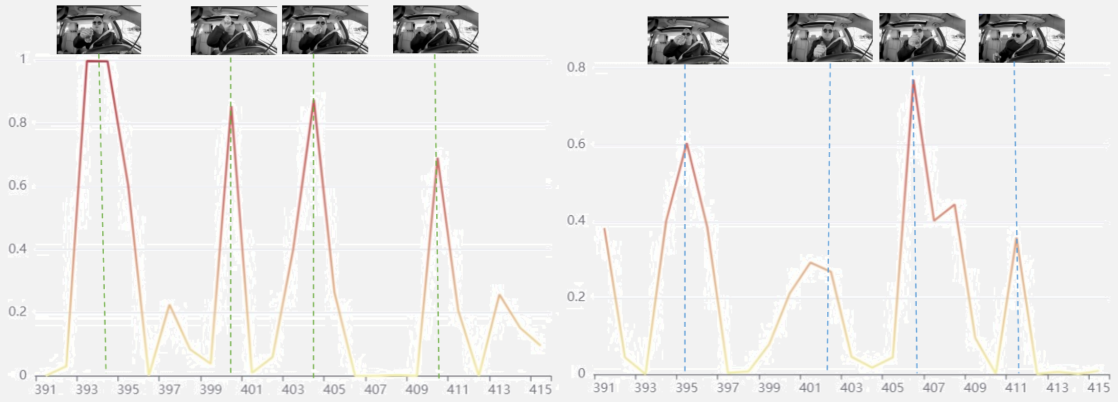
2

Methodology

3 Experimental Results

1 Background Settings

Localization Boundary Refinement



Start Signal Score   End Signal Score

Problem of refining boundaries only depending on the strongest signal.

Some points that do not represent starting or ending points are highly similar to actual starting or ending points in cyclical behaviors' signals. To address this problem, for cyclical behaviors, when we search for signals in the neighbor, we not only consider the intensity of signals, but also make judgments in combination with the number of times the signal appears.

2022 AI City Challenge

3 Experimental Results

1 Background Settings

2 Methodology

$$F1 = \frac{2TP}{2TP + FP + FN}$$

F1-Score	Precision	Recall
0.2902	0.4868	0.2067

Metrics & Visualization Results.

F1 determines the harmonic mean of precision and recall. A true-positive(TP) action identification will be considered when the action was correctly identified as starting time within one second and ending time within one second of the action.

Our model can accurately localize the action boundary due to the introduced of DBG-based Signal Acquisition module. We evaluate our methodology on the Track 3 validation data and obtain F1-Score at 0.2902, with a precision of 0.4868 and recall of 0.2067.

Thank you for your Listening!

2022 AI City Challenge

Presented by WHU-IIP