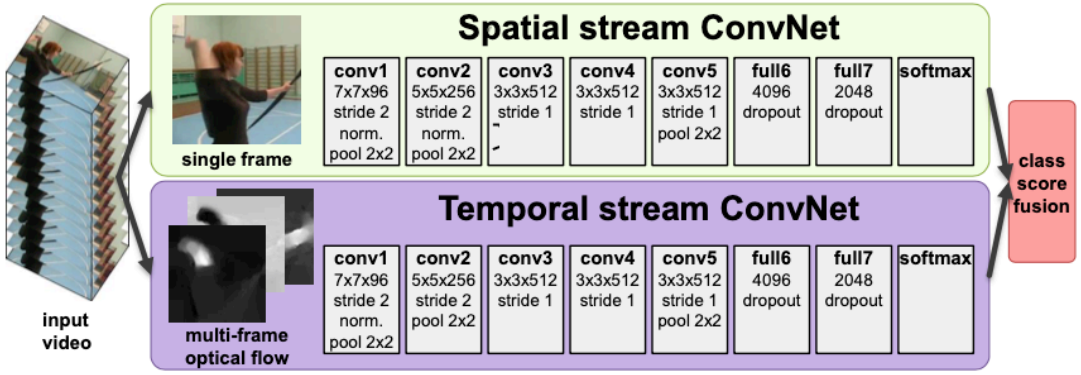
Pose-aware Multi-feature Fusion Network for Driver Distraction Recognition

Mingyan Wu[1, 2], Xi Zhang[1, 2], Linlin Shen[1, 2*], Hang Yu[3]
[1]Computer Vision Institute, College of Computer Science and Software Engineering
[2]Guangdong Key Laboratory of Intelligent Information Processing
[3]Shenzhen Laboratory of IC Design for Internet of Things
Shenzhen University, PR China
wumingyan2018@email.szu.edu.cn, zhangxi1987see@gmail.com, {llshen, yuhang}@szu.edu.cn

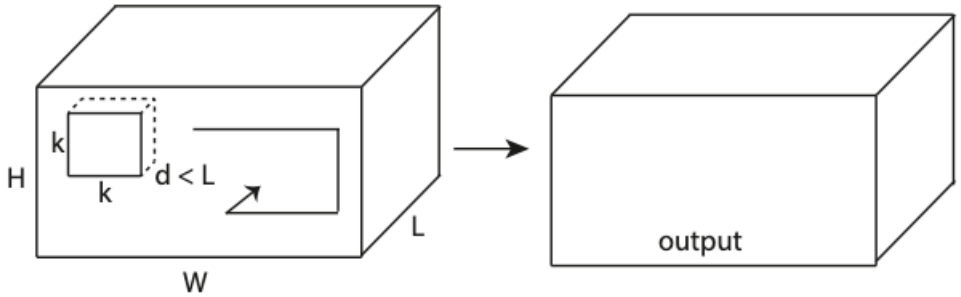
Introduction

- Distracted driving is an activity that diverts attention from driving. Distracted driving is one of the major causes of traffic accidents.

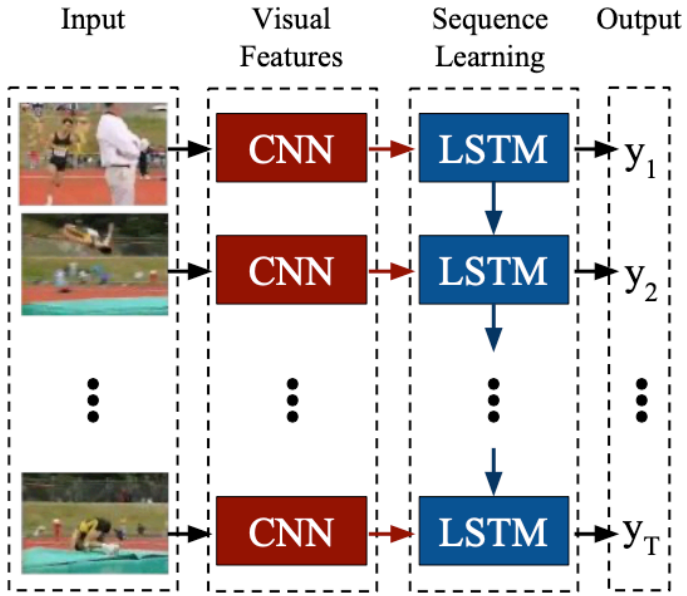
Related Work



Two-stream network



C3D: 3D convolutional network

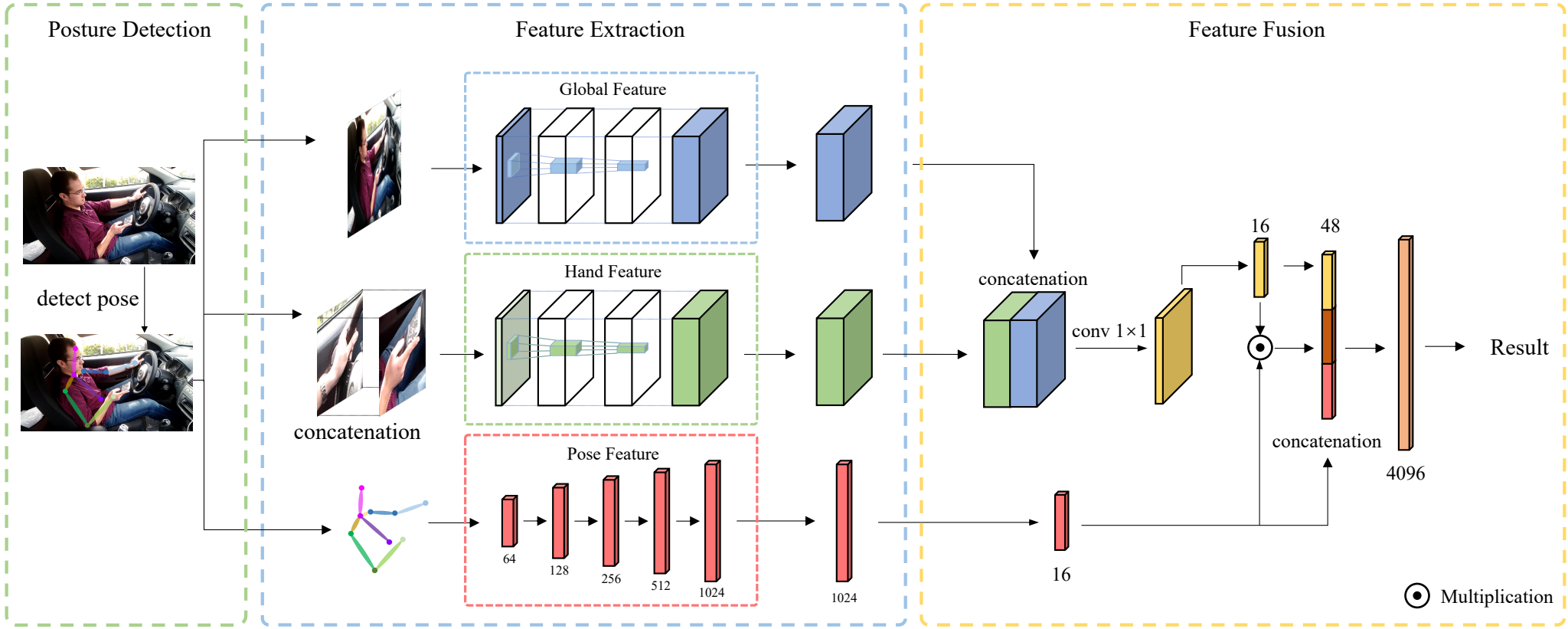


Long-term Recurrent Convolutional Networks

Method

- The whole original image contains global information.
- The actions of hand are important cues in driver distraction recognition.
- The pose information is robust against the interference of backgrounds.

Method


Posture Detection
Feature Extraction
Feature Fusion
detect pose
concatenation
Global Feature
Hand Feature
Pose Feature
64
128
256
512
1024
1024
concatenation
conv 1×1
16
48
concatenation
4096
Result
16
Multiplication

Datasets

- SZ Bus Driver Dataset.

TABLE I
DETAILS OF SZ BUS DRIVER DATASET

Class	Action	Training set	Test set
C0	Playing the phone	23973	5232
C1	Talking on the phone	5553	1495
C2	Smoking	14008	962
C3	Safe driving	13686	993
Total	Total	57220	8682

(a) C0: Playing the phone

(b) C1: Talking on the phone

(c) C2: Smoking

(d) C3: Safe driving

Datasets

- AUC Distracted Driver Dataset.

	Distracted Driver V1	Distracted Driver V2
Key Contributions	- First publicly available dataset for distracted driving - Training and testing datasets are split randomly	- Collected more data with more drivers - More precise labeling and better sampling per class - Training and testing datasets are split based on drivers
Dataset Information	31 drivers	44 drivers

C0: Safe Driving C1: Text Right C2: Phone Right C3: Text Left C4: Phone Left

C5: Adjusting Radio C6: Drinking C7: Reaching Behind C8: Hair or Makeup C9: Talking to Passenger

Experimental Results

TABLE II
RESULT ON SZ BUS DRIVER DATASET

Feature	Backbone	C0	C1	C2	C3	Total
Global	VGG-16	74.04	83.01	41.78	85.80	73.35
	ResNet-50	87.11	47.89	48.85	83.48	75.70
	InceptionV3	83.19	82.14	22.48	90.53	77.28
Late Fusion	VGG-16	90.84	84.55	72.25	92.45	88.78
	ResNet-50	93.12	92.78	75.36	91.14	90.87
	InceptionV3	95.85	90.63	80.87	92.74	92.93
Early Fusion	VGG-16	94.13	95.38	82.35	66.16	91.09
	ResNet-50	95.43	**99.46**	64.13	94.76	92.58
	InceptionV3	**96.46**	97.66	**89.40**	**95.27**	**95.75**

SZ Bus Driver Dataset.

TABLE III
RESULT ON AUC V1 AND V2 DATASET

Dataset	Method	Accuracy
AUC V1	GA-Weighted Ensemble (2017) [21]	95.98
	DenseNet+Latent Pose (2018) [23]	94.20
	VGG with Regularization (2018) [10]	**96.31**
	I3D-two stream (2019) [26]	77.10
	AlexNet+HOG features (2019) [7]	93.19
	Our method	96.28
AUC V2	GA-Weighted Ensemble (2019) [22]	90.07
	Our method	**90.38**

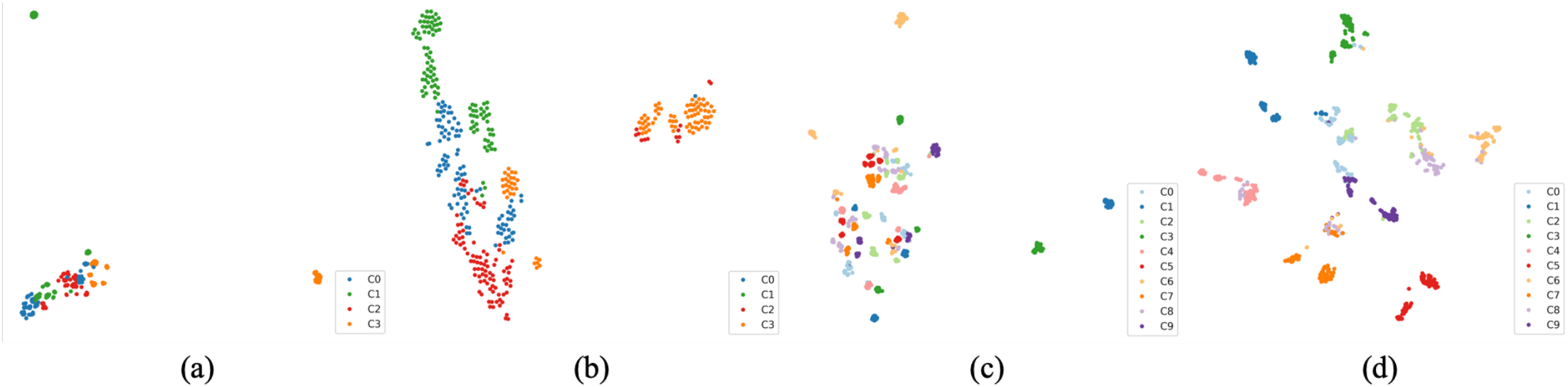
AUC Dataset.

Experimental Results

TABLE IV
ABLATION STUDY ON THREE DIFFERENT DATASETS

Dataset	Feature			Accuracy
	Global	Hand	Pose	
SZ Bus Driver	✓			77.28
		✓		85.58
			✓	88.29
		✓	✓	91.35
	✓	✓		88.68
	✓		✓	91.84
	✓	✓	✓	**95.75**
AUC V1	✓			95.22
		✓		90.86
			✓	91.36
		✓	✓	92.06
	✓	✓		95.65
	✓		✓	95.52
	✓	✓	✓	**96.28**
AUC V2	✓			85.12
		✓		67.86
			✓	74.88
		✓	✓	79.36
	✓	✓		87.15
	✓		✓	87.31
	✓	✓	✓	**90.38**

Visualization



The t-SNE visualization of global feature for SZ dataset (a) and AUC V2 dataset (c) and that of fused feature for SZ dataset (b) and AUC V2 dataset (d).

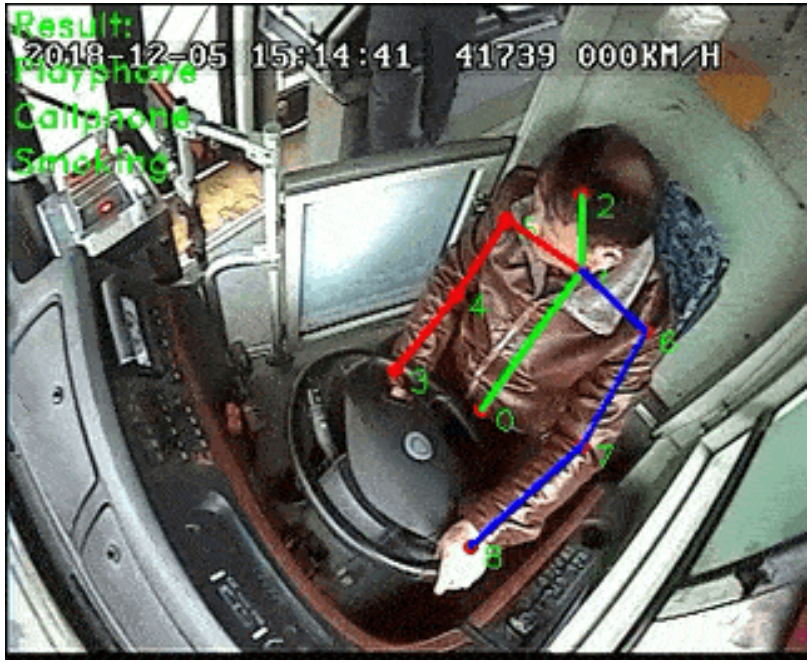
Visualization

Phone Left

Smoking

Safe Driving

Visualization



Thanks!