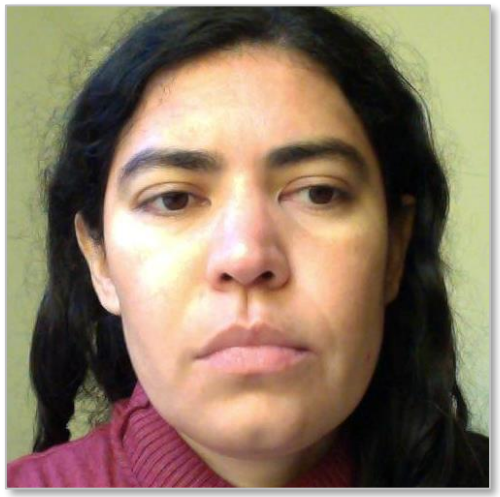
ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

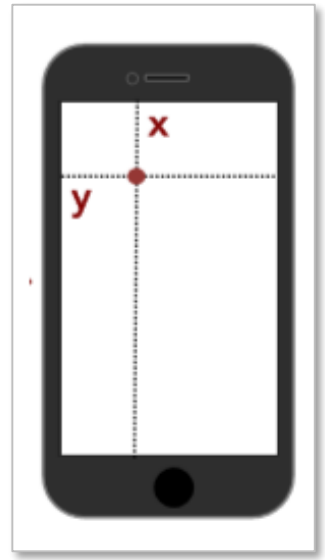
Adaptive Feature Fusion Network for Gaze Tracking in Mobile Tablets

Yiwei Bao[1], Yihua Cheng[1], Yunfei Liu[1], Feng Lu[1, 2]

[1]State Key Laboratory of Virtual Reality Technology and Systems, School of CSE, Beihang University, Beijing, China. [2]Peng Cheng Laboratory, Shenzhen, China

Introduction Gaze Estimation

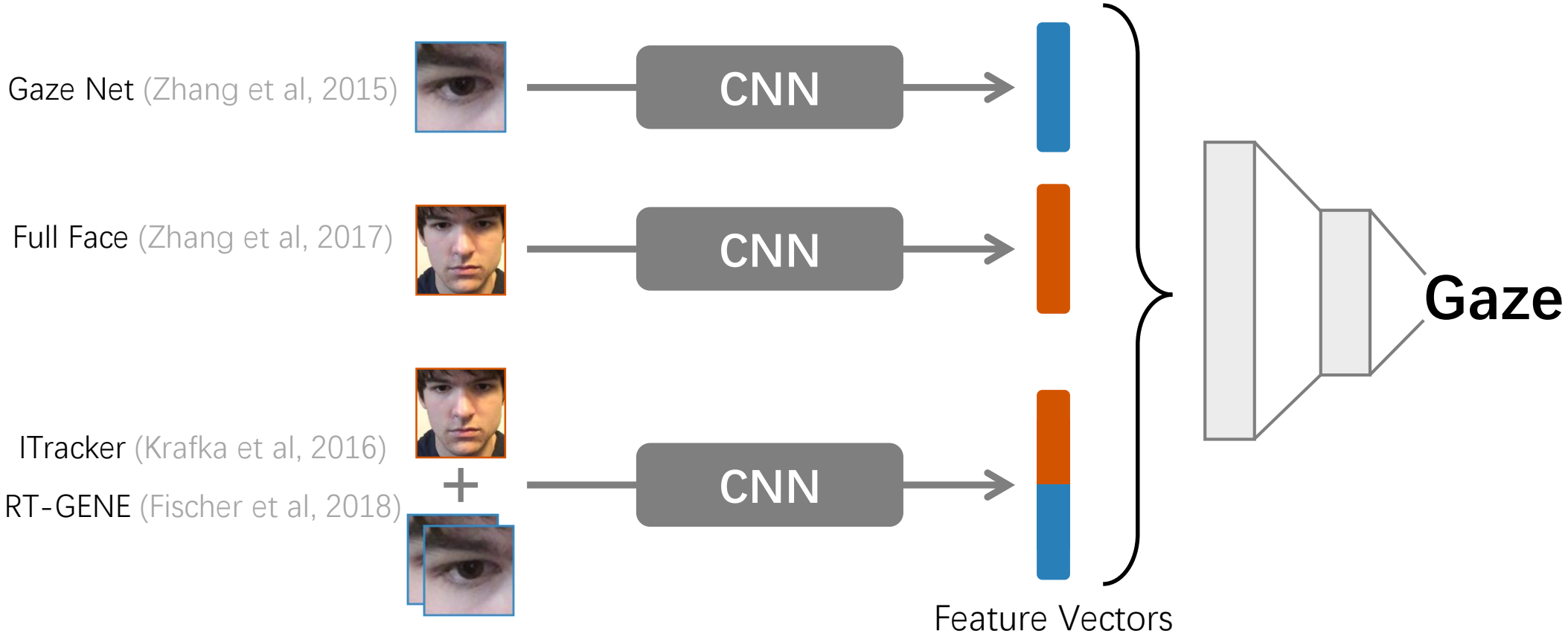




Facial image of the subject

2D gaze location on screen

Related Works

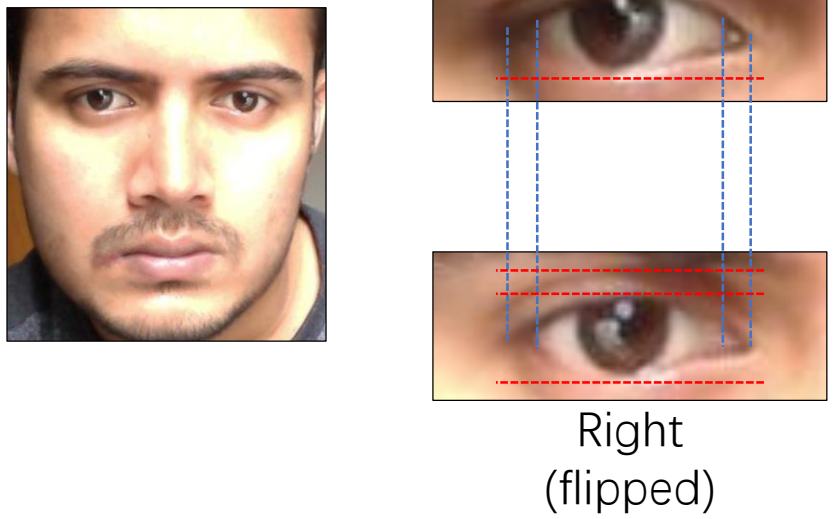


Limitation

1. Neglect the special similarity of two eye structures
2. Treat face and eye images separately.

Motivation

Two Eye Similarity



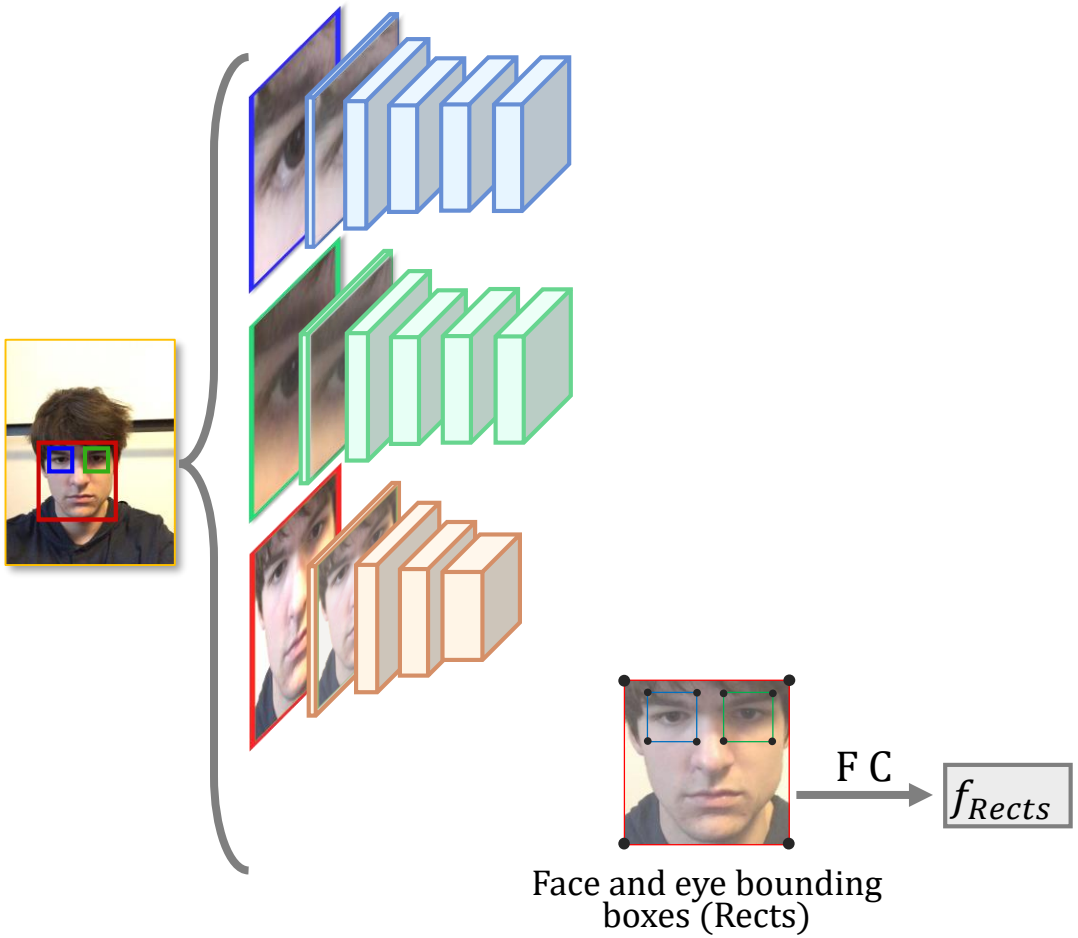
Structures like canthus, eyelid and outline of two eyes are almost identical.

Relationship between eye appearance and face

Eye appearance varies under different circumstances. Face appearance gives some clues (head pose, glasses, et al.) about how eyes would look like.

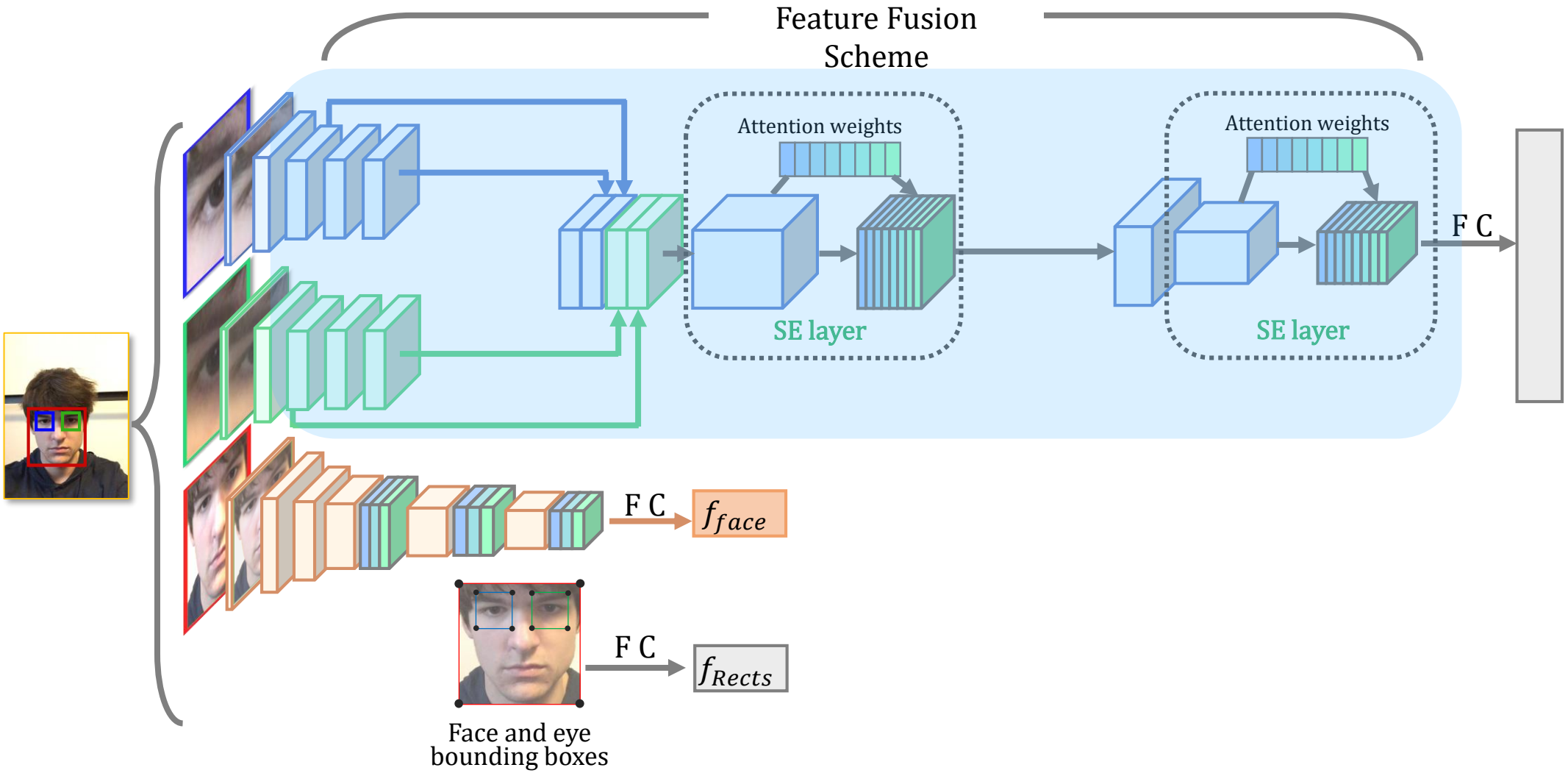
Method Adaptive Feature Fusion Network (AFF-Net)

Four stream inputs: eye images, face image, face and eye bounding boxes locations (Rects)



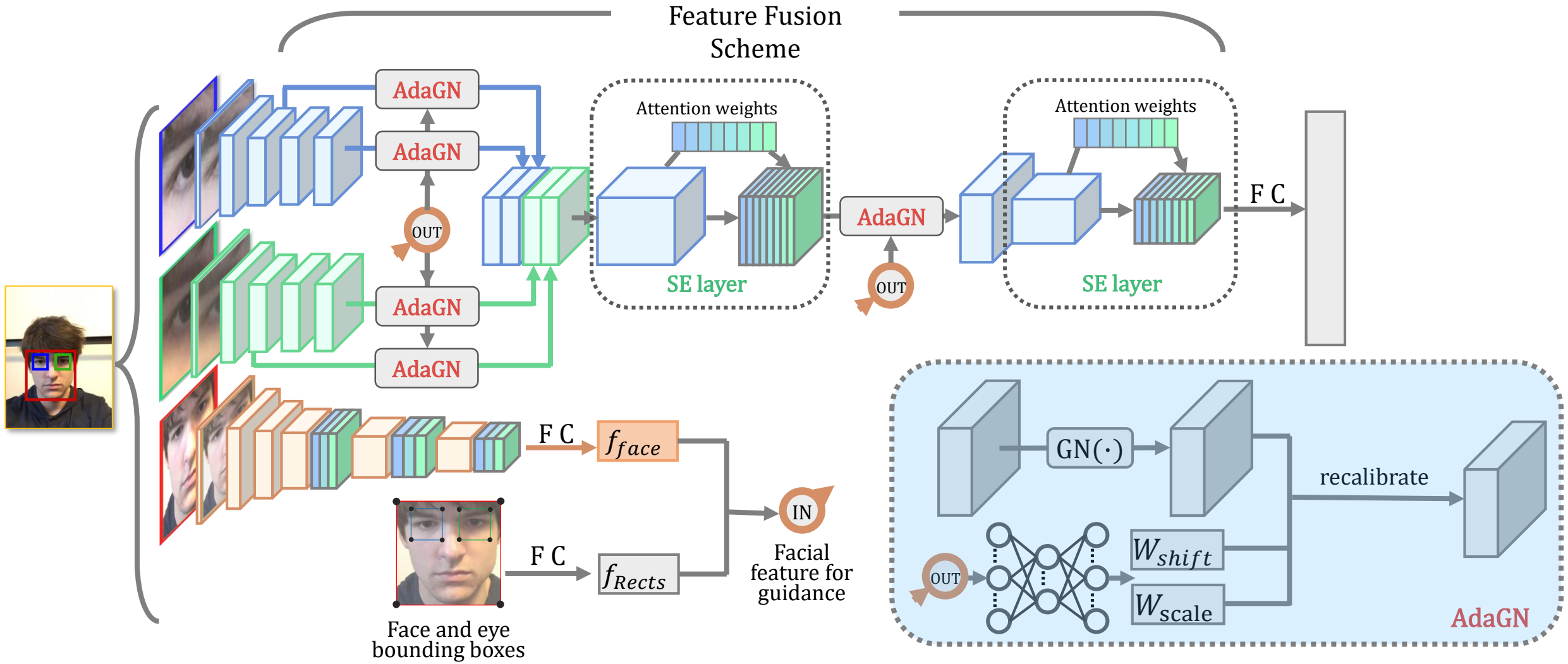
Method Adaptive Feature Fusion Network (AFF-Net)

Feature Fusion Scheme: eye feature fusion via stacking, attention weights and conv layers.

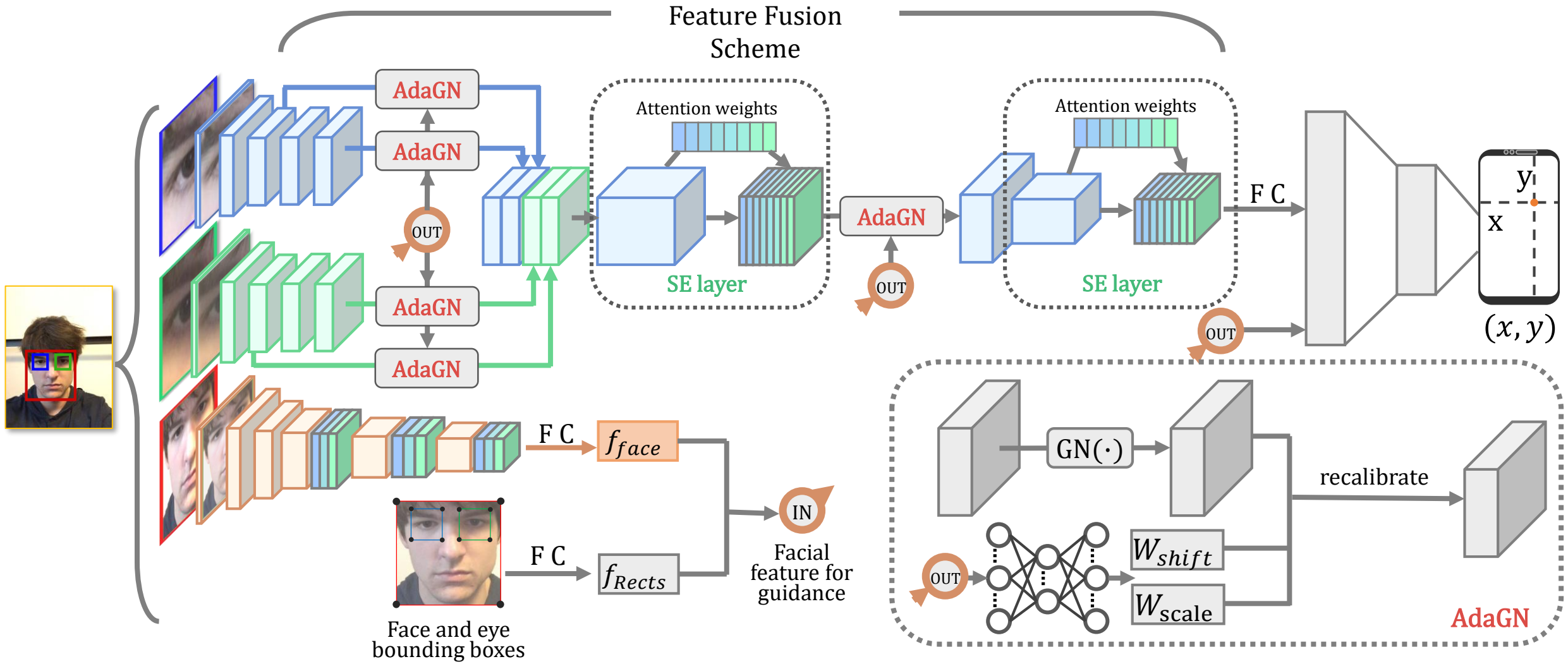


Method Adaptive Feature Fusion Network (AFF-Net)

Adaptive Group Normalization: recalibrate eye features according to facial features

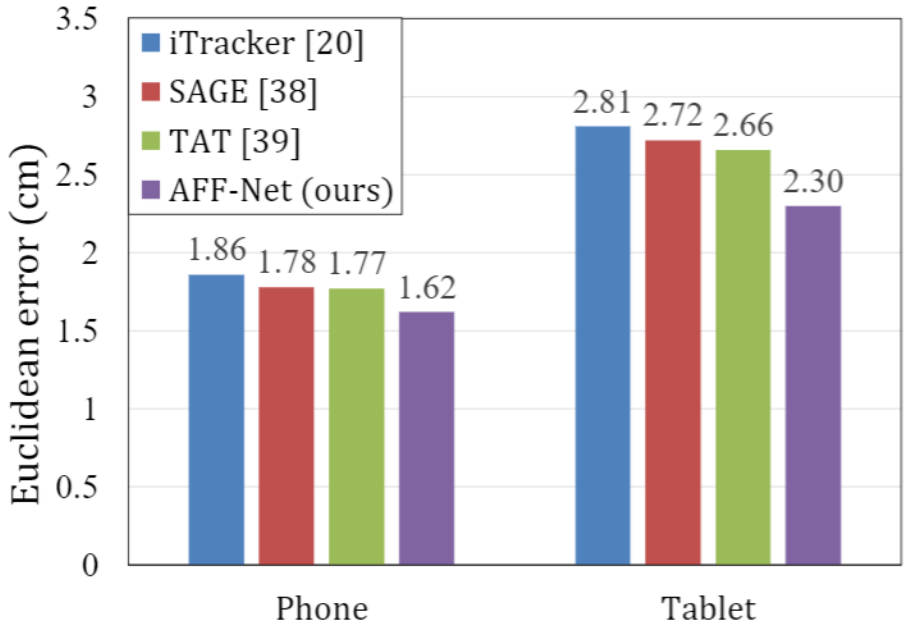


Method Adaptive Feature Fusion Network (AFF-Net)



Experiments GazeCapture Dataset

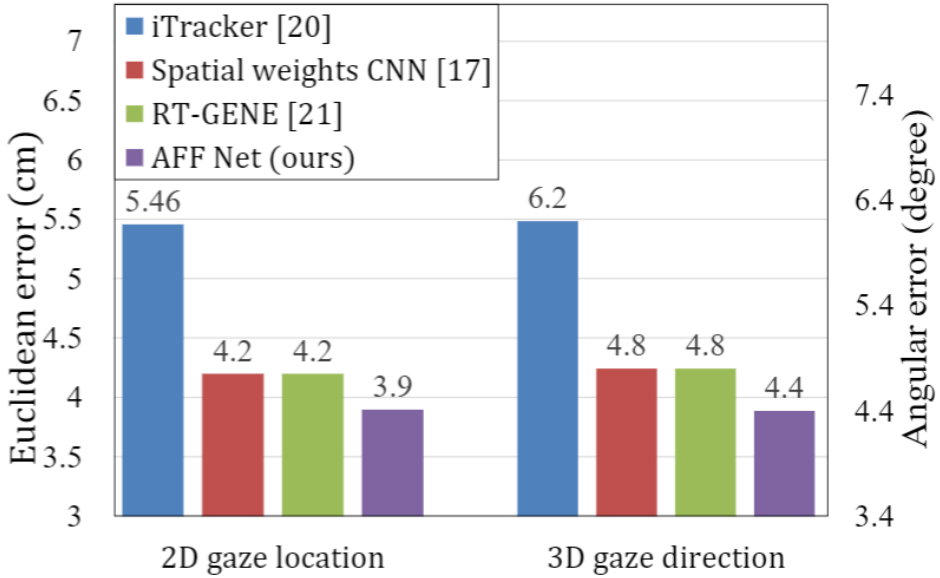
Method	Phone error (cm)	Tablet error (cm)
iTracker [20]	1.86	2.81
SAGE [40]	1.78	2.72
TAT [41]	1.77	2.66
AFF-Net	**1.62**	**2.30**



Output performs current SOTA methods on mobiles and tablets.

Experiments MPIIGaze Dataset

Method	2D gaze location error (cm)	3D gaze direction error (degree)
iTracker [20]	5.46	6.2
Spatial weights CNN [16]	4.2	4.8
RT-GENE [22]	4.2	4.8
AFF-Net	**3.9**	**4.4**



Also achieves SOTA performance on MPIIGaze dataset.

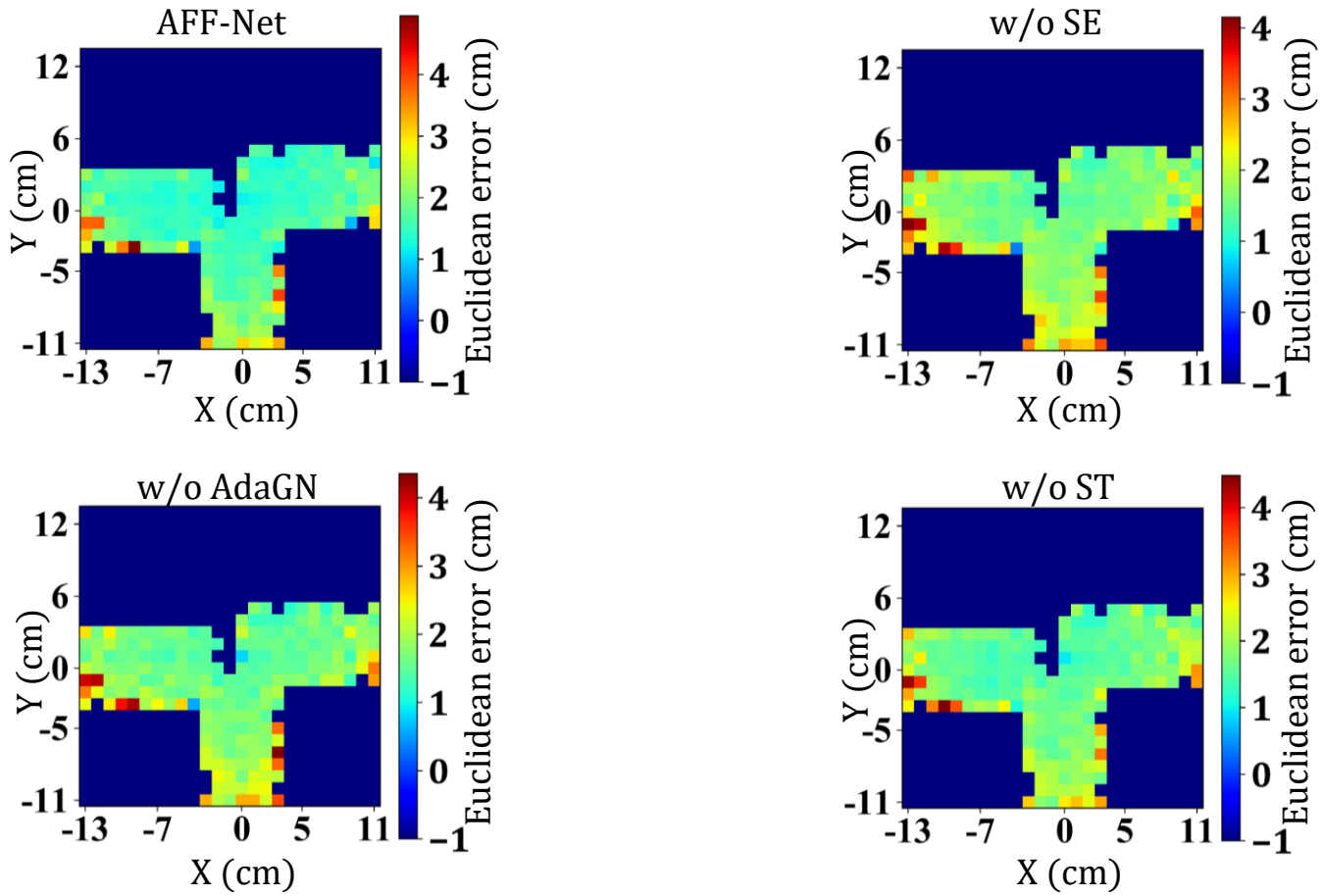
Experiments

Ablation study——GazeCapture dataset

Method	GazeCapture	
	Phone (cm)	Tablet (cm)
AFF-Net	**1.62**	**2.30**
without ST	1.67	2.39
without SE	1.68	2.31
without AdaGN	1.69	2.33

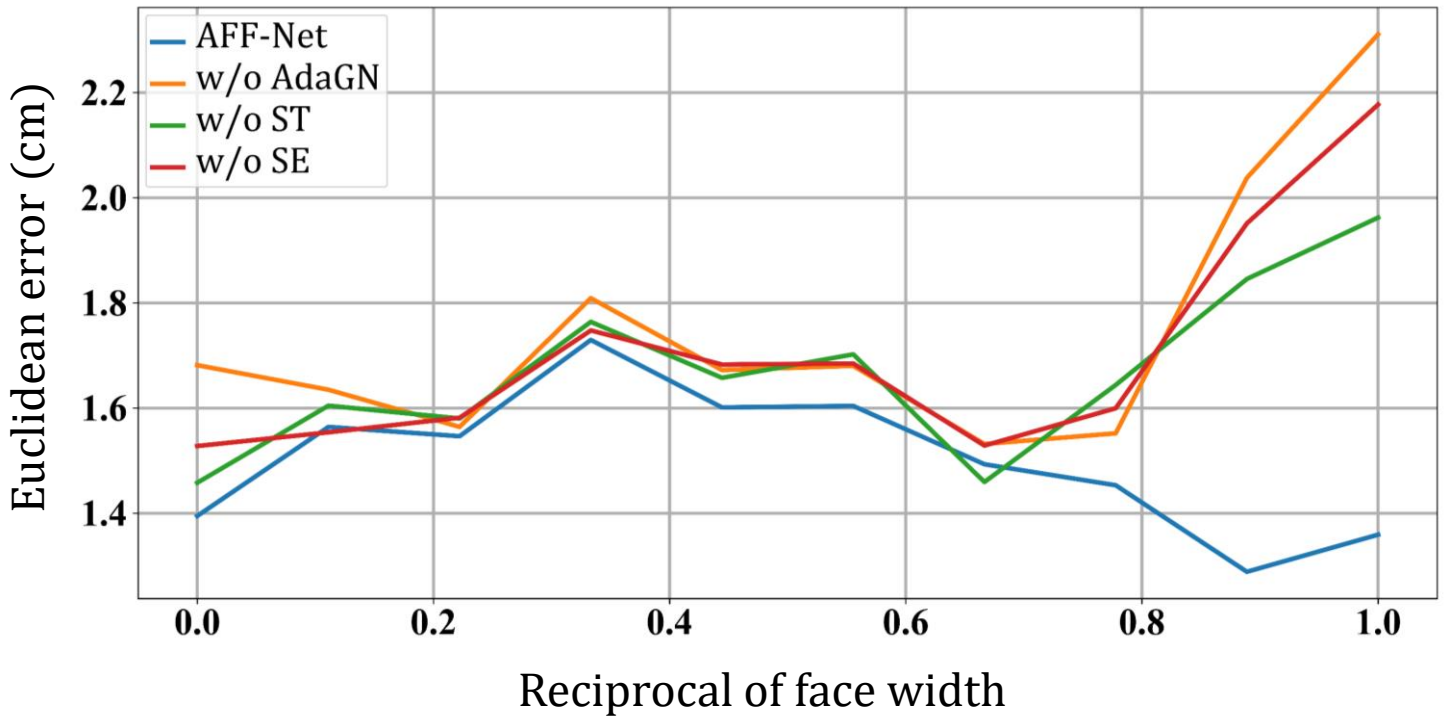
Ablation study proves effectiveness of each module.

Experiments Ablation study——GazeCapture dataset



Further analysis shows that proposed modules reduce estimation errors. Especially on those **remote locations.**

Experiments Ablation study——GazeCapture dataset



Proposed network gives more accurate estimation **when eyes and face are small in the image.**

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Thanks

lufeng@buaa.edu.cn

http: // phi-ai.org