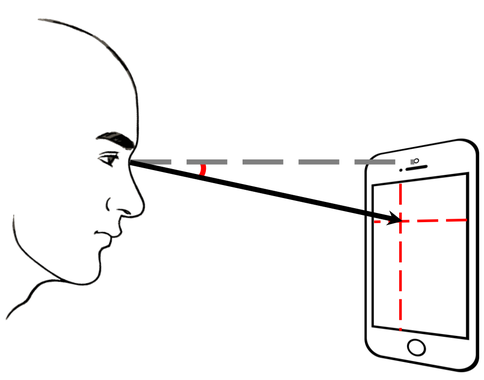
Adaptive Feature Fusion Network for Gaze Tracking in Mobile Tablets

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.

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

Introduction – Gaze Estimation



Definition

Infer subject's gaze direction or gaze location from the subject's frontal facial images.

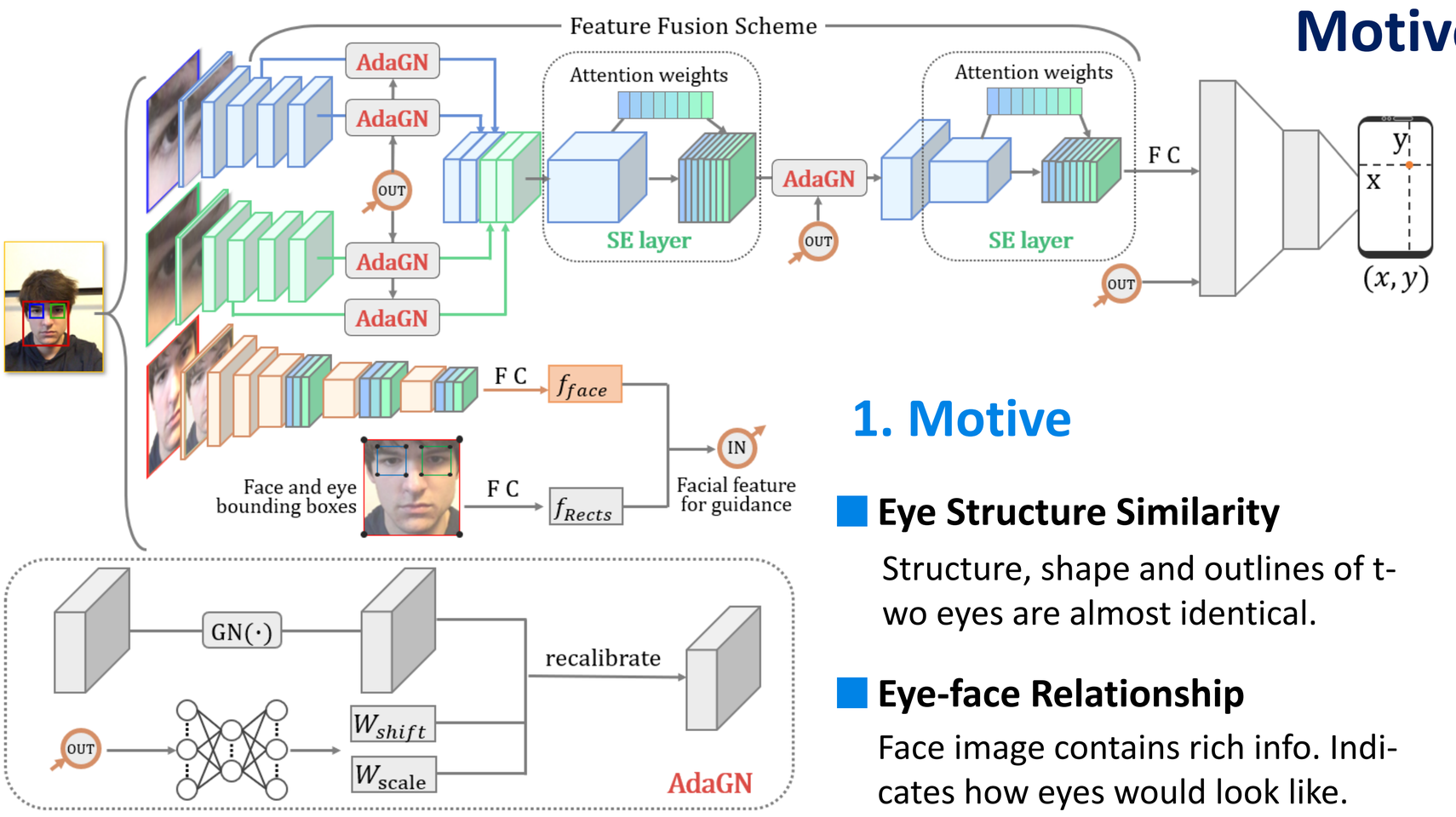
Eye Gaze

Important indicator of human attention and cognition.
Reveals how people interact w-ith their surroundings.

Application

Mental condition diagnose and hum-an intention prediction.
Newly-developing human com-puter interaction method.

Motive and Methods



1. Motive

Eye Structure Similarity

Structure, shape and outlines of t-wo eyes are almost identical.

Eye-face Relationship

Face image contains rich info. Indi-cates how eyes would look like.

2. Methods

Adaptive Feature Fusion Network (AFF-Net)

Four stream inputs:
Eye images, face image, face an-d eye bounding boxes locations.

Feature Fusion Scheme

Feature fusion by stacking, att-ention weights and conv layers.

Adaptive GN

Adaptive eye feature recalibrat-ion according to facial features.

Experimental Results

Performance on Public Datasets

Methods	GazeCapture		MPIIGaze	
	Phone (cm)	Tablet (cm)	2D (cm)	3D (degree)
iTracker	1.86	2.81	5.46	6.2
SAGE	1.78	2.72	4.2	4.8
TAT	1.77	2.66	4.2	4.8
Ours	**1.62**	**2.30**	**3.9**	**4.4**

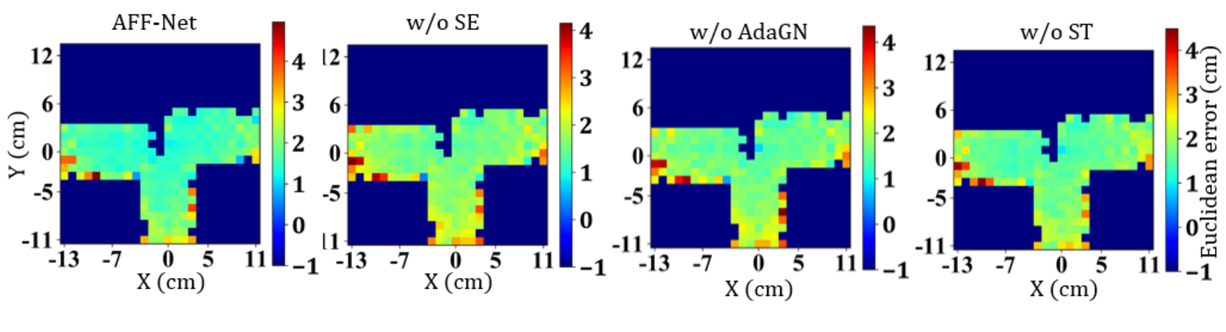
AFF-Net outperforms state-of-the-art methods on widely used GazeCapture and MPIIGaze dataset.

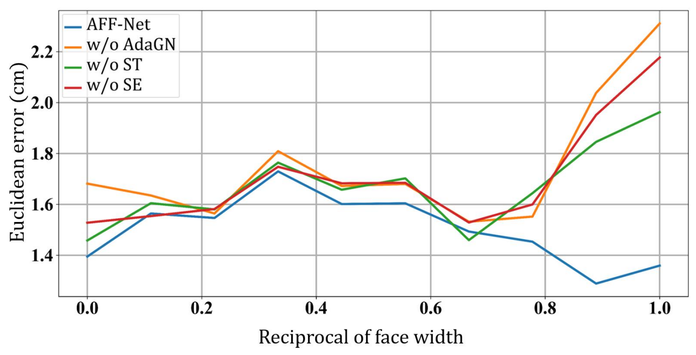
Ablation Study

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

Ablation study shows the effectiveness of each module.

Result Analysis





Conclusion

AFF-Net performs better o-n hard cases: remote locat-ions and small faces.

Case Study