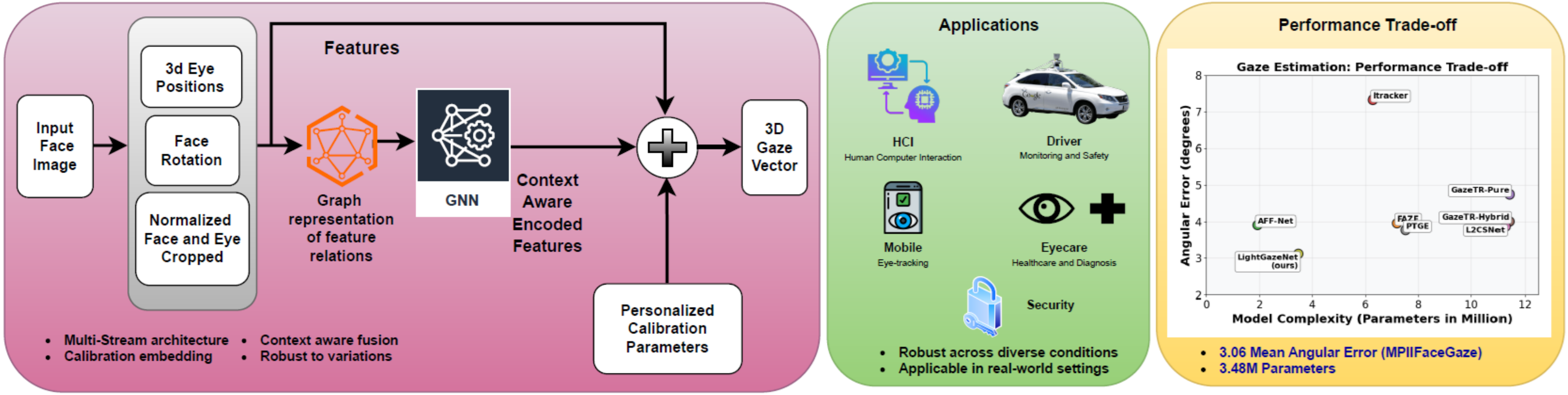
LightGazeNet: A Lightweight GNN-based Architecture for Gaze Estimation

Heena Patel, Anirban Chowdhury, Pooja Jigar Choksy,
Samiksha Pradeep Pachade, Ajinkya Puar

Akeso Eyecare, Beijing, China

Introduction

- Gaze tracking determines where a person is looking using facial and eye information
- Critical technology for human-computer interaction and behavior understanding
- Growing demand for lightweight, accurate solutions for edge devices
- Applications include healthcare diagnostics, automotive safety systems, educational engagement tracking, mental health assessment, etc.

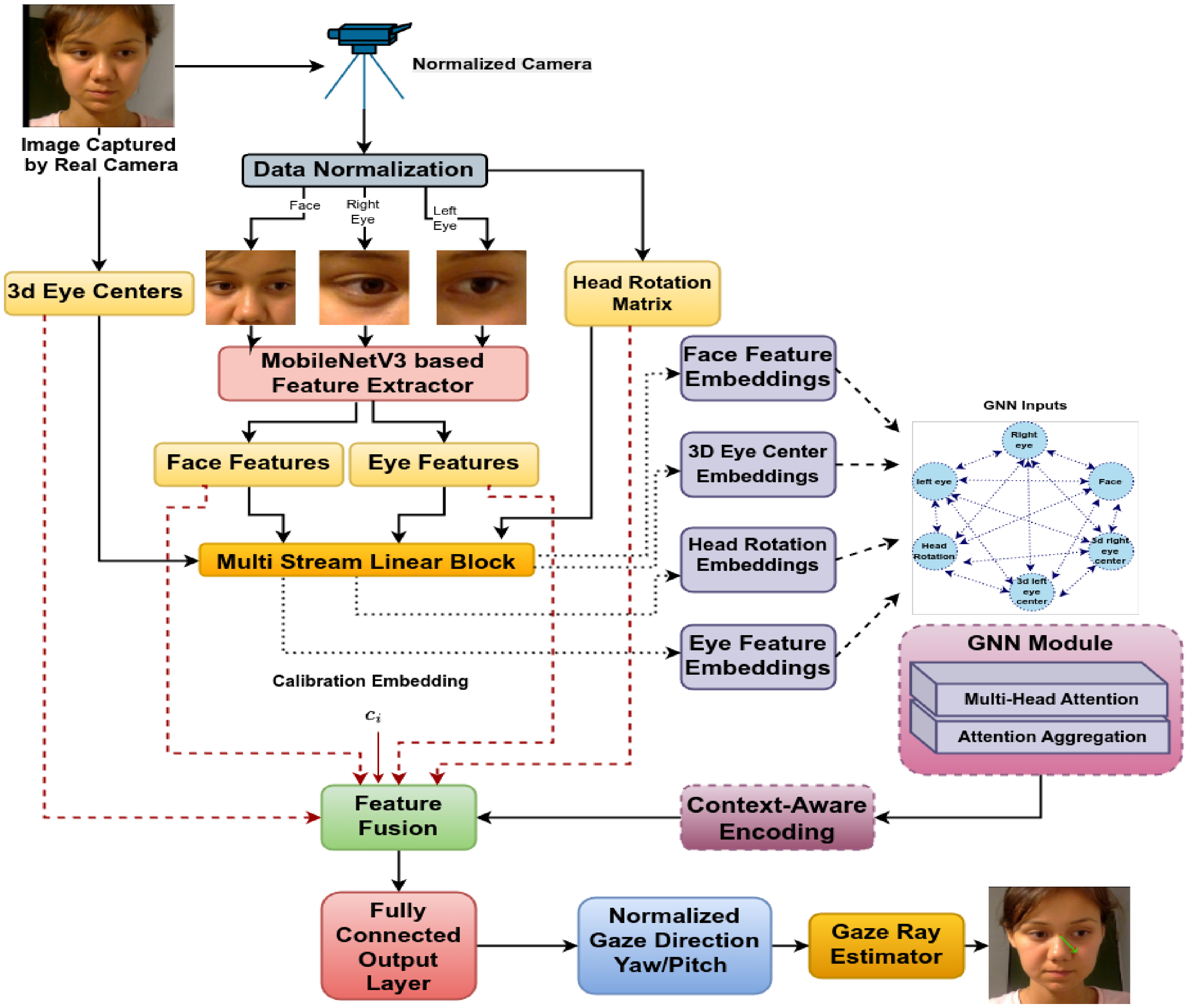


Research Gap

- Need for explicit modeling of spatial and semantic relationships between multi-modal gaze features
- Lack of lightweight architectures that maintain high accuracy
- Missing: effective feature fusion that leverages complementary information across modalities

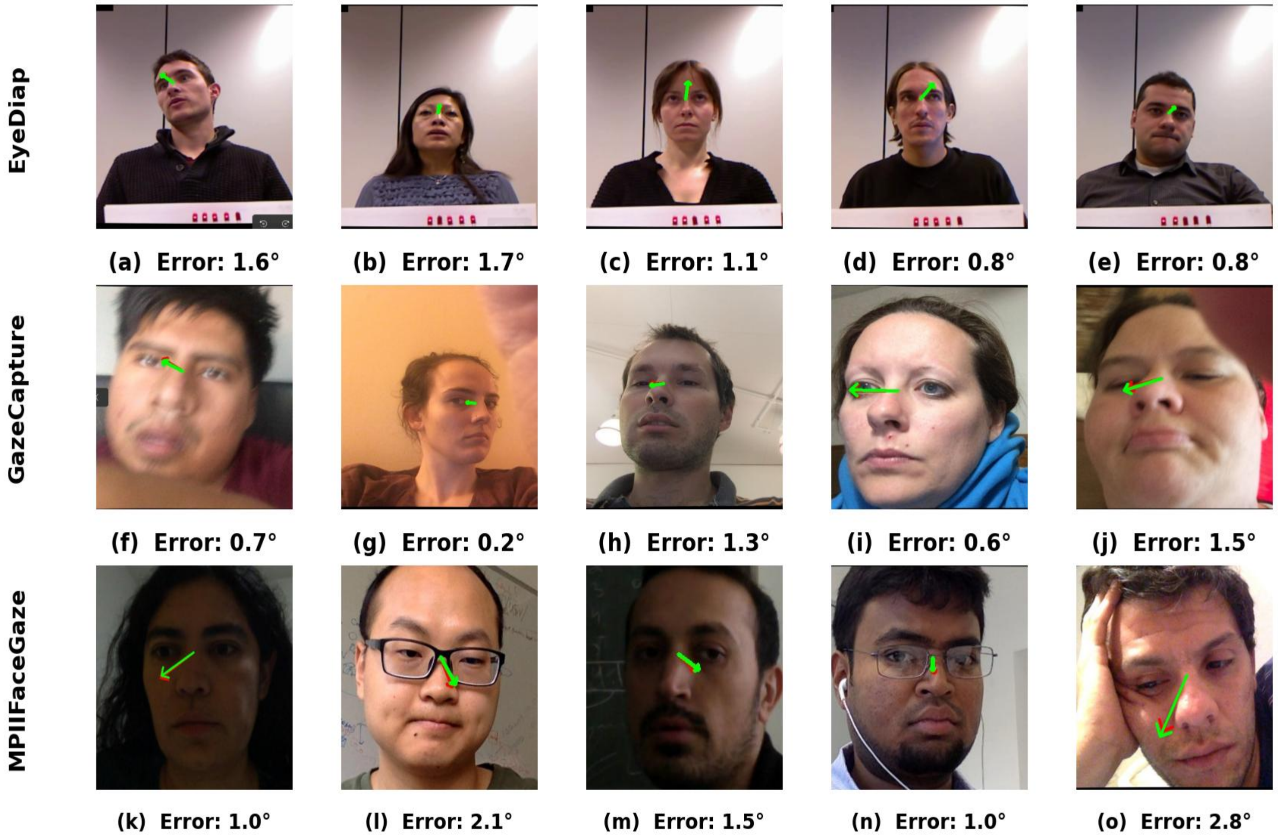
Key Contributions

- Novel application of Graph Neural Networks to gaze estimation
- Multi-head Attention: 4-head attention mechanism for cross-modal reasoning
- Lightweight architecture enabling real-time inference on edge devices
- Real-world Deployment: Real-time inference on edge devices



LightGazeNet architecture overview. The framework processes heterogeneous inputs through lightweight encoders, constructs a fully connected graph, applies multi-head attention-based GNN reasoning, and outputs gaze predictions via regression.

Results: Benchmark Comparison



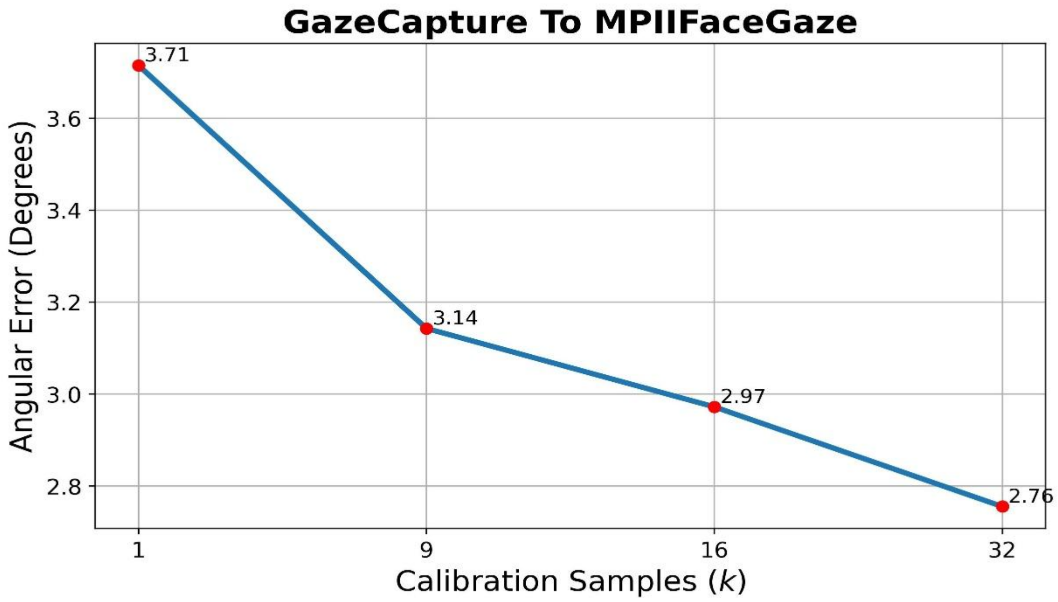
Qualitative results showing gaze prediction under visual uncertainties such as blur, occlusion, and head pose variation. Samples (a–e) are from EyeDiap; (f–j) from GazeCapture and (k -o) MPIIFaceGaze.

SOTA Accuracy on 3 benchmarks

Method	MPIIFaceGaze (°)	EyeDiap (°)	GazeCapture (cm)
Itracker	7.33	7.13	2.34
FAZE	3.95	4.31	1.73
AFF-Net	3.90	6.41	1.96
GazeTR-Hybrid	4.00	5.17	2.31
PTGE	3.76	3.34	1.88
L2CSNet	3.86	3.05	1.77
LightGazeNet	**3.06**	**2.91**	**1.69**

- MPIIFaceGaze: LightGazeNet 3.06° vs PTGE 3.76° vs L2CSNet 3.86°
- EyeDiap: LightGazeNet 2.91° vs L2CSNet 3.05° vs PTGE 3.34°
- GazeCapture: LightGazeNet 1.69cm vs FAZE 1.73cm vs L2CSNet 1.77cm

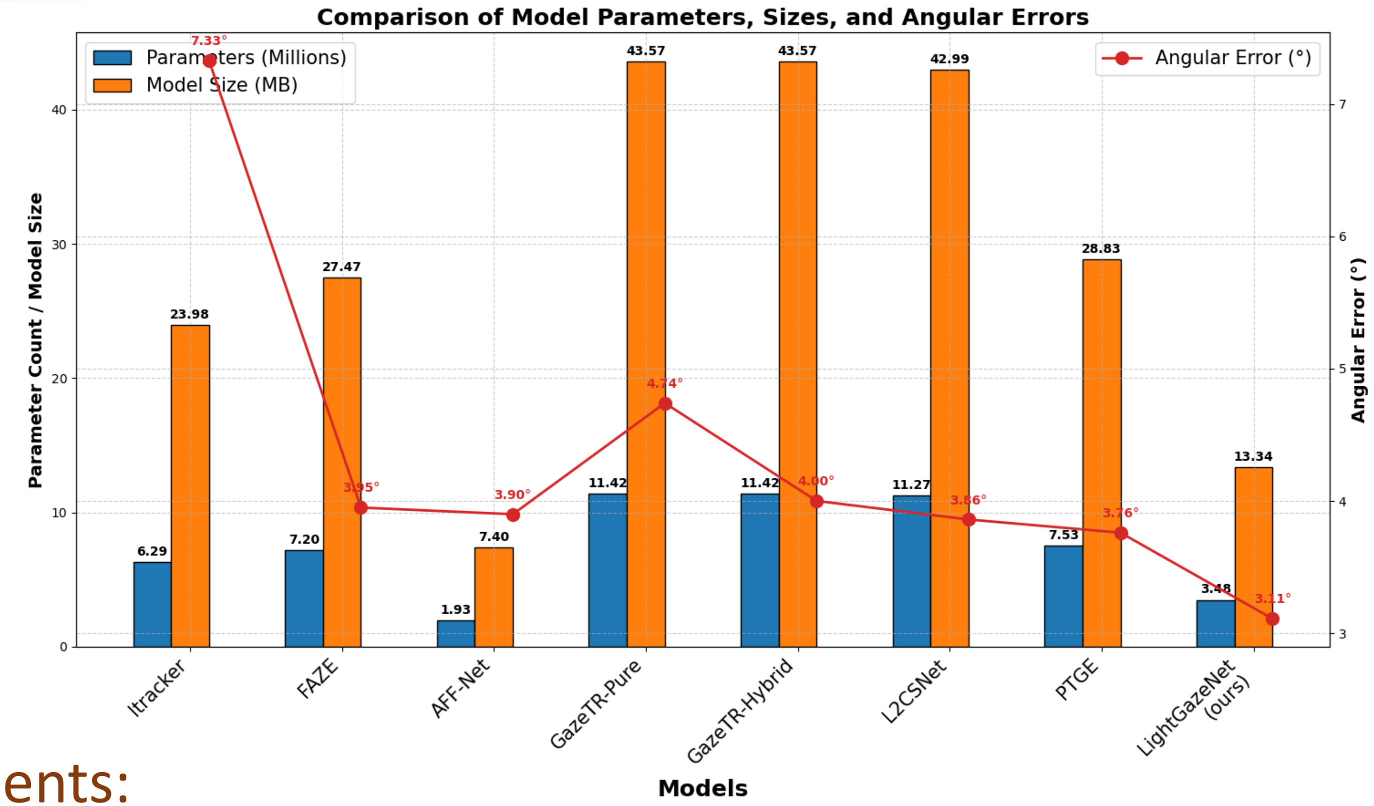
Cross-Dataset Generalization



Cross-Dataset Evaluation (Train: GazeCapture → Test: MPIIFaceGaze):

- Zero-shot transfer: 3.71° angular error
- With 9 calibration samples: 3.14° (15.4% improvement)
- With 16 calibration samples: 2.97° (19.9% improvement)
- With 32 calibration samples: 2.76° (25.7% improvement)
- Efficient adaptation with minimal target domain data demonstrates strong generalization capability of graph-based feature fusion.

Model Parameters vs Accuracy



Key Achievements:

- SOTA accuracy on all 3 benchmarks
- 54% parameter reduction vs PTGE (3.48M vs 7.53M)
- 18.6% accuracy improvement over PTGE
- Real-time inference: 18.65 MB model size

Conclusions

- Novel GNN-based gaze estimation with multi-head attention achieving SOTA on 3 benchmarks
- 3.06° error with only 3.48M parameters (54% reduction vs PTGE)
- 25.7% error reduction in cross-dataset generalization
- Real-time capable on resource-constrained devices

Future Directions:

- Temporal sequence modeling for video-based gaze tracking
- Self-supervised domain adaptation for enhanced robustness

AKESO

Thank You…!

For more information, you can reach us at eyelignai@akesoeyecare.com,
or visit our project website.