

## ABSTRACT

### FEATURE-BASED EYE TRACKING FOR HANDHELD DEVICES USING MOBILE-OPTIMIZED CONVOLUTIONAL NEURAL NETWORK

Submitted by:

Albert Cristianto Halim

22/492268/PA/21099

Eye-tracking is a vital technology in human-computer interaction, enabling hands-free control and accessibility features on mobile devices. However, existing models often suffer from high computational latency, poor generalizability, and dependence on controlled environments. This research proposes a lightweight, multi-branch eye-tracking system optimized for handheld devices using MobileNetV3. The system performs gaze estimation through a three-input architecture processing full face images, dedicated eye crops ( $64 \times 64$  pixels with 20% padding), and geometric facial landmarks extracted via MediaPipe.

Unlike previous studies limited to controlled testing environments, this research processes portrait-oriented mobile captures under uncontrolled lighting conditions. The model employs a two-phase training strategy (frozen-to-unfrozen backbone) with Huber loss for outlier robustness, complemented by dropout regularization (0.25–0.40) and learning rate warmup. Performance is evaluated using Euclidean Mean Absolute Error (MAE) with percentile analysis.

Experimental evaluation on the GazeCapture dataset demonstrates that the proposed multi-branch architecture achieves 23.81% Mean Absolute Error (MAE) without calibration, improving to 19.67% MAE after per-user polynomial calibration a 17.4% relative improvement. In physical terms, this corresponds to reducing prediction error from 2.89 cm to 2.39 cm on a typical smartphone screen. The calibrated performance is competitive with lightweight approaches such as GazeHFR (22.06% MAE), while the model maintains only 5.3 million parameters suitable for mobile deployment.

**Key Words:** Eye Tracking, Gaze Estimation, Neural Network, Feature Based Tracking, Mobile Net, MediaPipe