BRNO UNIVERSITY OF TECHNOLOGY FACULTY OF INFORMATION TECHNOLOGY

Excel@FIT 2026
http://excel.fit.vutbr.cz

PupillaRX: On-Device Pupillary Reflex Analysis from Smartphone Video Recording

Alex Marinica*

Abstract
This work presents PupillaRX, a smartphone-first pipeline for pupillary reflex recording, quality control, and metric extraction without cloud processing. The method combines a manually annotated eye dataset and a lightweight MobileNet-style multi-head CNN that predicts pupil/iris geometry, blink state, and eye side jointly. On the latest 5-fold grouped cross-validation run, we reach ratio error between 0.048 and 0.060, pupil center error in the 2.6–2.9 px range, and iris center error in the 3.1–3.4 px range. Eye-side accuracy is in the range 0.885–0.908. The key contribution is not a new deep architecture, but a confidence-aware, end-to-end workflow that ties data annotation, real-time capture guidance, and on-device analysis into one practical research system.

*xmarina00@vutbr.cz, *Faculty of Information Technology, Brno University of Technology*

1. Introduction

Motivation. Pupillary light reflex (PLR) is clinically useful, but robust measurement typically needs dedicated hardware. Recent work shows that phone-based pupillometry is feasible [1, 2, 3], yet deployment still fails when recordings are low-quality (blink, reflections, poor alignment) or when analysis depends on cloud infrastructure.

Problem. For a student-scale system, we need a model and application that can run fully on device, estimate pupil/iris geometry frame-by-frame, reject unreliable recordings, and still output meaningful PLR metrics from real user captures.

Our solution. PupillaRX is an end-to-end workflow: (1) confidence-aware annotation of smartphone torch recordings, (2) lightweight multi-head CNN training, and (3) iOS capture-to-analysis integration in the Pupillary Reflex MVP app.

Main contribution. The contribution is integration and engineering rigor, not a new neural architecture: a confidence-aware, on-device pupillometry pipeline that couples live quality scoring with post-capture metric gating.

2. Dataset and Annotation

Data were recorded with a smartphone camera and torch to match the deployment setting. In the latest raw export we have 2,438 annotated tasks; 1,984 samples are retained for cross-validation after confidence and preprocessing filters.

Each frame annotation contains:

- Pupil ellipse, pupil confidence (`high`/`medium`/`low`), and pupil flags (`occluded`, `glint`, `blink`).
- Iris ellipse, iris confidence (`high`/`medium`/`low`), and iris flag (`occluded`).
- Eye side label (`left` or `right`).

This schema is directly reflected in training targets and in downstream quality heuristics, which reduces the gap between annotation logic and deployment behavior.

3. Model and Training

The model is a MobileNet-style depthwise-separable CNN backbone [4] with four heads:

- 2-channel heatmap head for pupil and iris centers (64×64),
- radii head for pupil/iris ellipse radii,
- blink classification head,
- eye-side classification head.

Preset	Ratio	Pupil (px)	Iris (px)	Eye acc.
Balanced	0.0559	2.79	3.08	0.892
Ratio-prio	0.0480	2.87	3.38	0.908
Center-prio	0.0599	2.64	3.10	0.885

Table 1. Mean 5-fold CV results on batch 2026_04_18 (1,984 samples).

Training uses a multi-task objective:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{hm}}\,\mathcal{L}_{\text{hm}} + \lambda_{\text{radii}}\,\mathcal{L}_{\text{radii}} + \lambda_{\text{blink}}\,\mathcal{L}_{\text{blink}} + \lambda_{\text{ratio}}\,\mathcal{L}_{\text{ratio}} + \lambda_{\text{eye}}\,\mathcal{L}_{\text{eye}}.$$

The heatmap term uses MSE; the radii and ratio terms use SmoothL1; and the blink and eye-side terms use BCE. Across runs, $(\lambda_{\text{hm}}, \lambda_{\text{radii}}, \lambda_{\text{blink}}) = (1.0, 6.0, 0.5)$, while λ_{ratio} and λ_{eye} are preset-dependent (Table 1); the ratio-priority run uses $(\lambda_{\text{ratio}}, \lambda_{\text{eye}}) = (1.4, 0.15)$.

With the current 7-layer feature setting, the network has 70,984 trainable parameters. Training uses grouped 5-fold CV by `eye_video_id` to avoid near-duplicate leakage between train and validation clips. Checkpoint selection uses a geometry-oriented score combining ratio error and center errors.

Eye-side accuracy is currently in the $0.885-0.908$ range, limited by the number of eye-side labeled samples. Blink-head accuracy is intentionally not reported as a headline metric because the current set contains only a small number of blink-positive samples. This is acceptable for UI auto-switch support, but further labeling is needed for robust clinical interpretation.

4. On-Device Application Pipeline

The iOS application integrates capture, analysis, interpretation, and persistence without network dependency:

1. Guided recording with baseline, torch-on, and post-torch phases.
2. On-device CoreML inference on every N-th raw frame during capture, feeding a smoothed live quality verdict.
3. Post-capture CoreML analysis over extracted frames.
4. PLR metric computation only when recording quality is sufficient.
5. Local persistence of test records (JSON).

Quality control is a first-class subsystem. Frame-level reasons include low heatmap confidence, blink-like frames, missing pupil/iris, excessive center offset, unstable ratio, blur, and saturation. Recording-level score combines global usable-frame fraction and baseline stability:

$$\text{score} = 0.45\,u_{\text{frame}} + 0.55\,u_{\text{baseline}} - \text{penalties}.$$

Metrics are blocked for score < 0.70, which aligns computation with recording quality and prevents false confidence from poor captures.

5. Conclusions

PupillaRX demonstrates that clinically motivated pupillometry can be run fully on-device with practical robustness safeguards. The strongest current value is integration quality: annotation schema, multi-task model outputs, and quality-aware application behavior are consistent end-to-end.

At the same time, the current results are not sufficient to claim replacement of standard clinical pupillometry. Additional validation is required before such claims for specific use cases, including paired recordings against reference clinical pupillometers, broader testing across devices and lighting conditions, and prospective studies on larger and more diverse cohorts.

Acknowledgements

I would like to thank my supervisor, prof. Ing. Adam Herout, Ph.D., for his guidance in the ML pipeline and for practical tips during data collection. I also thank doc. Ing. Radim Kolář, Ph.D., prof. MUDr. Evžen Růžička, DrSc., FCMA, FEAN, and MUDr. Olga Bartošová, Ph.D., for their guidance and support throughout the project. I also thank everyone who helped collect eye-reflex data for training purposes; this work would not have been possible without their help.

References

[1] Colin Barry, Jessica de Souza, Yinan Xuan, Jason Holden, Eric L. Granholm, and Edward Jay Wang. At-home pupillometry using smartphone facial identification cameras. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, pages 1–12, New York, NY, USA, 2022. Association for Computing Machinery.

[2] Omar Solyman, Mokhtar Mohamed Ibrahim Abushanab, Andrew R. Carey, and Amanda D. Henderson. Pilot study of smartphone infrared pupillography and pupillometry. *Clinical Ophthalmology*, 16:303–310, 2022.

[3] Raffaele Mazziotti, Fabio Carrara, Aurelia Viglione, Leonardo Lupori, Luca Lo Verde, Alessandro Benedetto, Giulia Ricci, Giulia Sagona, Giuseppe Amato, and Tommaso Pizzorusso. Meye: Web app for translational and real-time pupillometry. *eNeuro*, 8(5):ENEURO.0122–21.2021, 2021.

[4] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural net-

works for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.