# Neuromorphic Eye-Tracking for Low-Latency Pupil Detection


Paul Hueber[1], Luca Peres[1], Florian Pitters[2], Alejandro Gloriani[2], Oliver Rhodes[1]

1: University of Manchester, UK

2: Viewpoint System, Austria

MANCHESTER 1824
The University of Manchester

# Eye-Tracking

- Eye Tracking allows for understanding a user's state and intent and has applications for medical and psychological diagnostics, augmented and virtual realty (AR / VR), robotics and driver monitoring systems

- Our eyes are fastest-moving organ, with saccades of $15^\circ$ amplitude, and velocities exceeding $500^\circ/\mathrm{s}$.

- The latency of conventional cameras is bound by the frame-rate and kHz cameras require much energy, limiting types of eye—movements that can be monitored.

- Typically, model-based or appearance-based approaches are computationally intensive or lack predictive accuracy.

https://viewpointsystem.com/

https://www.nimbleai.eu/

VIEWPOINTSYSTEM

nimble AI

# Pupil-Tracking with 3ET+ Dataset

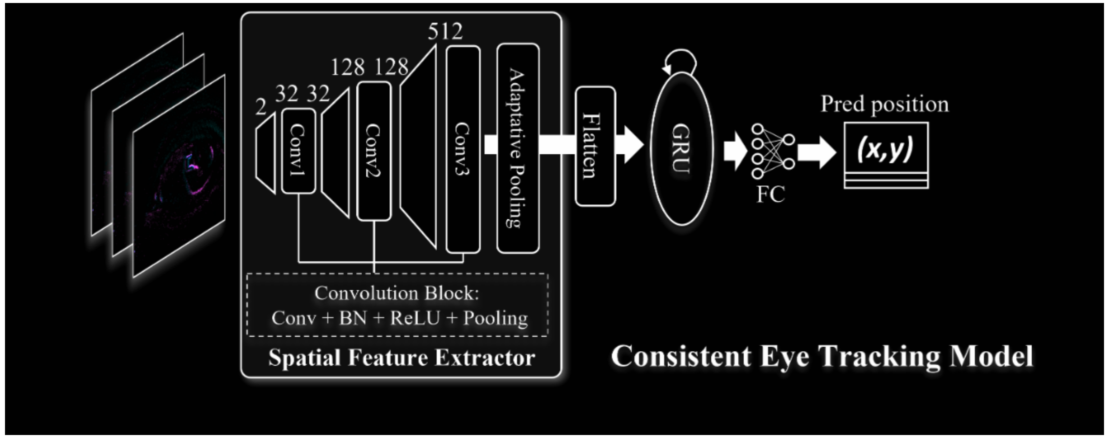
- 3ET+ event-based eye-tracking dataset:
  - 13 participants, multiple sessions,
  - DVXplorer Mini event camera
  - Resolution 640x480, 5MHz
- Behaviours: random eye movements, saccades, reading, smooth pursuit, blinks.
- Labels: pupil centre at 100 Hz plus blink indicator; evaluation via mean pixel error and PX-accuracy at various radii.
- AIS Challenge 2024 (Wang, 2024) and subsequent 2025 Workshop. Winning solutions are accurate, but high computational cost and latency.

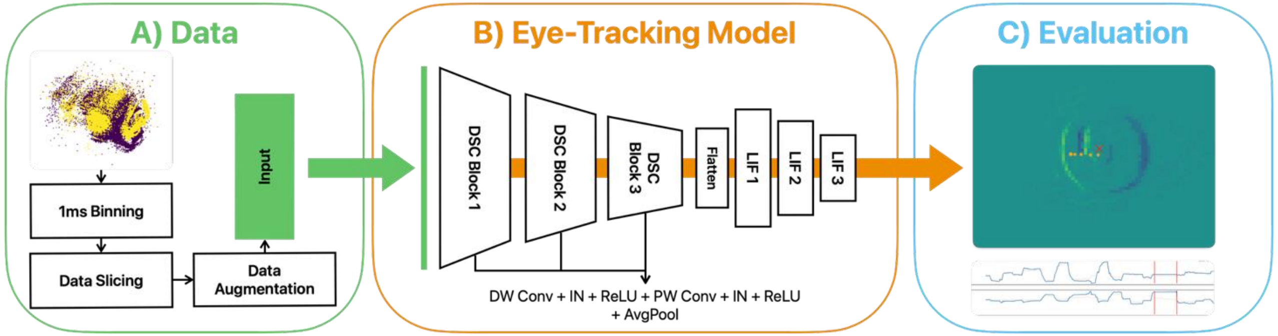
| Model | P1 | P10 | Euc. Dist. | Params | FLOPs |
|---|---|---|---|---|---|
| GTechVision [21] | 4.16 | 92.26 | 4.94 | 410k | |
| CB-ConvLSTM [34] | | 84.80 | 3.82 | 417k | 1.09T |
| PEPNet [35] | 7.79 | 97.95 | 3.51 | 640k | 459M |
| GoSparse [21] | 7.32 | 99.00 | 3.51 | 465k | |
| MeMo [21] | 6.53 | 99.05 | 3.20 | 5.4M | 230M |
| CETM [21] | 23.91 | 99.26 | 2.03 | 7.1M | 2.9G |
| ERVT [21] | 28.80 | 98.21 | 1.98 | 150k | 157M |
| MambaPupil [36] | 33.75 | 99.42 | 1.67 | 8.59M | 2.61T |
| TDTracker [37] | | 99.20† | 1.50 | 3.04M | 265M |
| BigBrain [38] | 45.50 | 99.00 | 1.44 | 809k | 110.4M |
| BRAT [39] | | **99.59†** | **1.14** | 7.1M | 5.8G |

Wang, Z., et al. "*Event-based eye tracking. ais 2024 challenge survey*." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024.

# Model Architecture

- Prior neuromorphic eye trackers (e.g. Retina) achieve ≈3.24 px error at <5 mW, but with accuracy or continuous-tracking limitations.
- Start with backbone from top AIS 2024 eye-tracking models with ConvLSTMs, transformers, or state-space modules.
- Systematically substitute temporal modules with stacked LIF layers that preserve temporal modelling capacity.
- Introduce depth-wise separable convolutions to cut parameters by ~6× and operations by ~82× in these backbones.

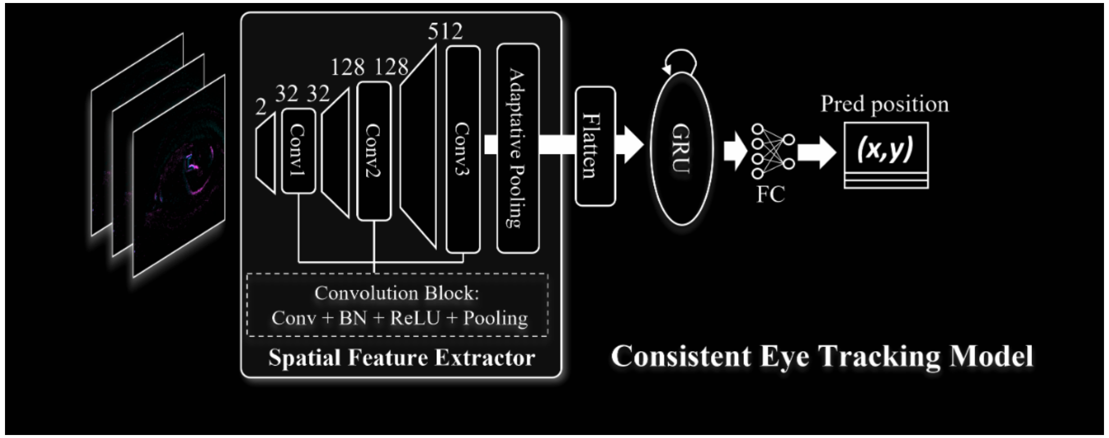


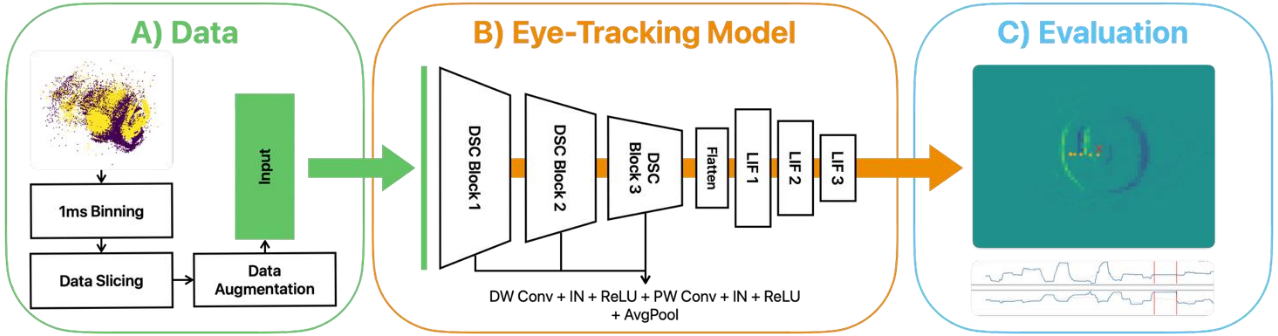




Wang, Z., et al. "*Event-based eye tracking. ais 2024 challenge survey*." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024.

# Experimental Setup

- Event stream binned into 1 ms slices; 450 ms windows used during training with augmentation (spatial, temporal, CutOut).
- Data down-sampled to 80x60 (to match competitors), and labels up-sampled to 1kHz using b-spline interpolation
- Best model selected on validation, then run in a continuous online setting to simulate real deployment.
- Loss: Positional MSE + velocity MSE, with inference evaluated according to PX-Accuracy



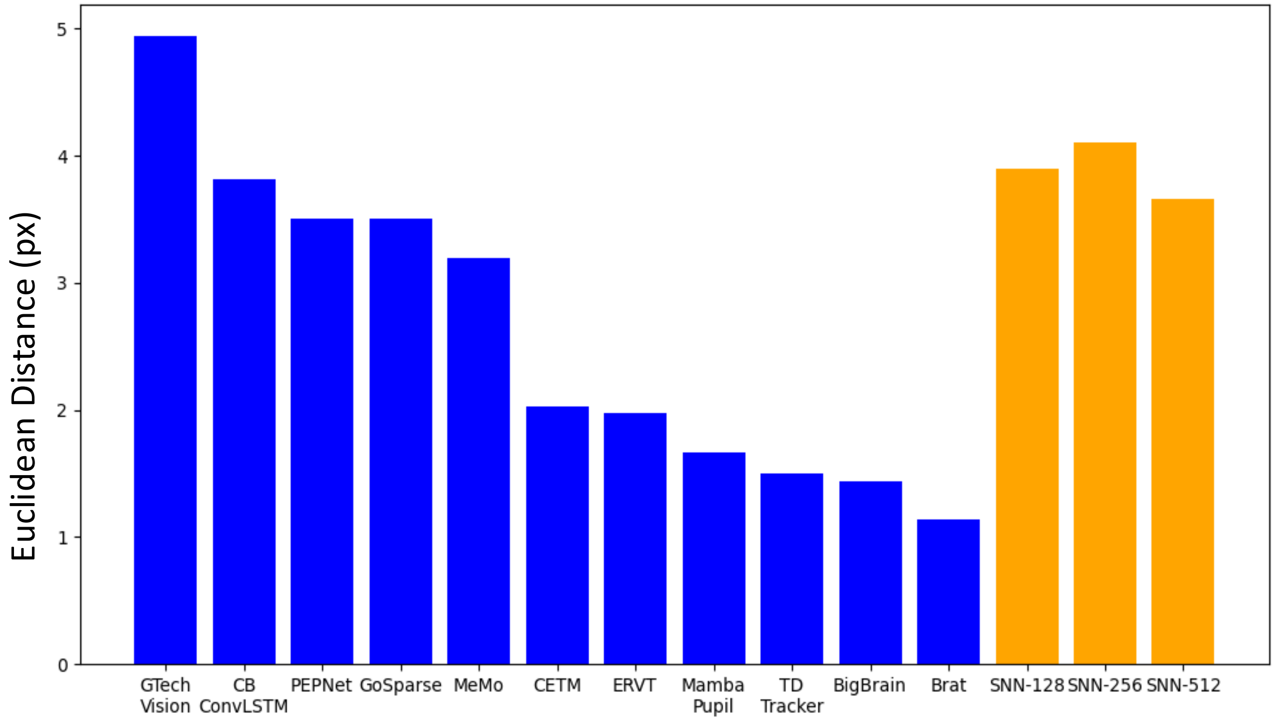

$$L = L_{Pos} + L_{Vel}$$

$$L_{Pos} = \sqrt{(x_{pred} - x_{gt})^2 + (y_{pred} - y_{gt})^2}$$

$$L_{Vel} = \sqrt{(\Delta_t x_{pred} - \Delta_t x_{gt})^2 + (\Delta_t y_{pred} - \Delta_t y_{gt})^2}$$

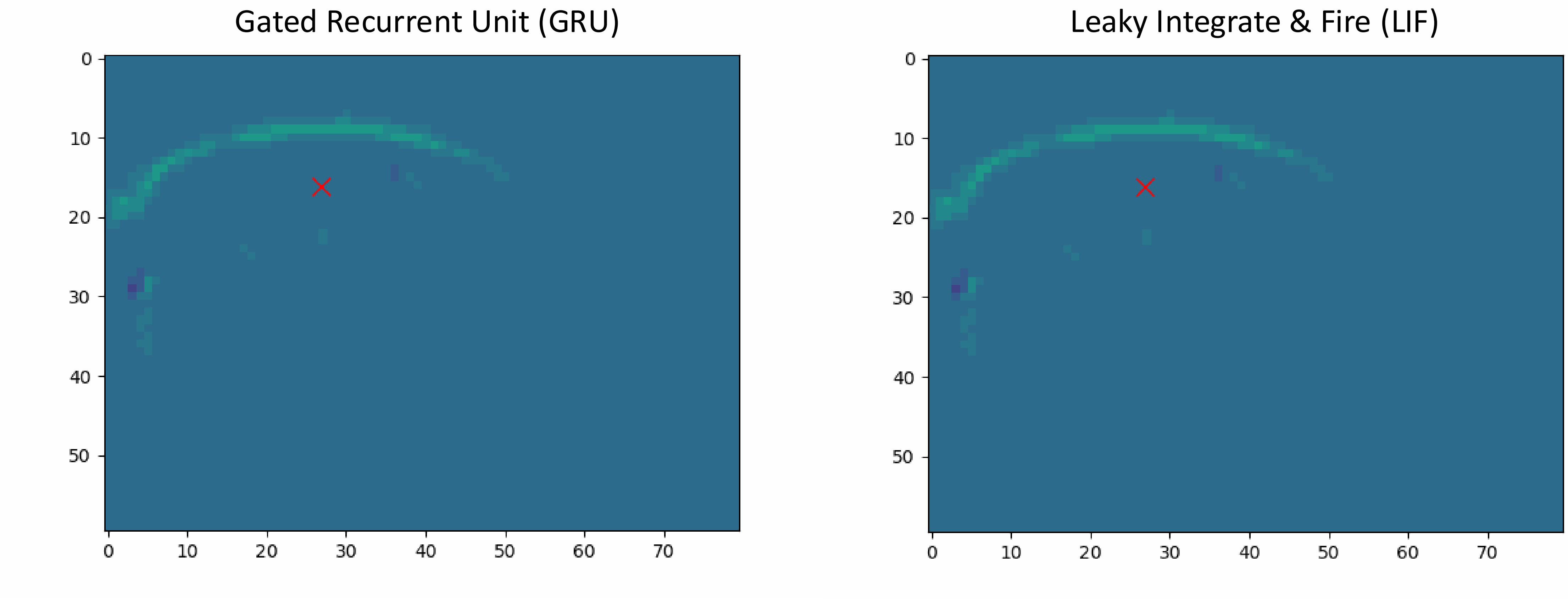
# Accuracy vs State-of-the-Art

- ANN baselines (e.g. BRAT, BigBrain) reach ≈1.1–1.5 px mean error but at high compute and 20 Hz evaluation.
- Proposed SNN variants achieve 3.7–4.1 px mean error on 3ET+, close to neuromorphic Retina's 3.24 px.
- Accuracy remains suitable for real-time pupil detection in wearable settings.
- Retina remains the main 'neuromorphic' reference solution (3.24 px, 2.89–4.8 mW, 5.57–8.01 ms), with our SNNs reaching comparable error (3.7–4.1 px) in similar power and latency regime




| Variant | Params | FLOPs | Power | Frequency | Latency |
|---|---|---|---|---|---|
| Retina [19] | 63k | 6.06M | <5 mW[2] | ≤5 kHz | >5 ms[2] |
| Our (N=512) | 353k | 3.4M | 4.9 mW[3] | 1 kHz | 3 ms |
| Our (N=256) | 189k | 3.1M | 4.2 mW[3] | | |
| Our (N=128) | 107k | 2.9M | 3.9 mW[3] | | |

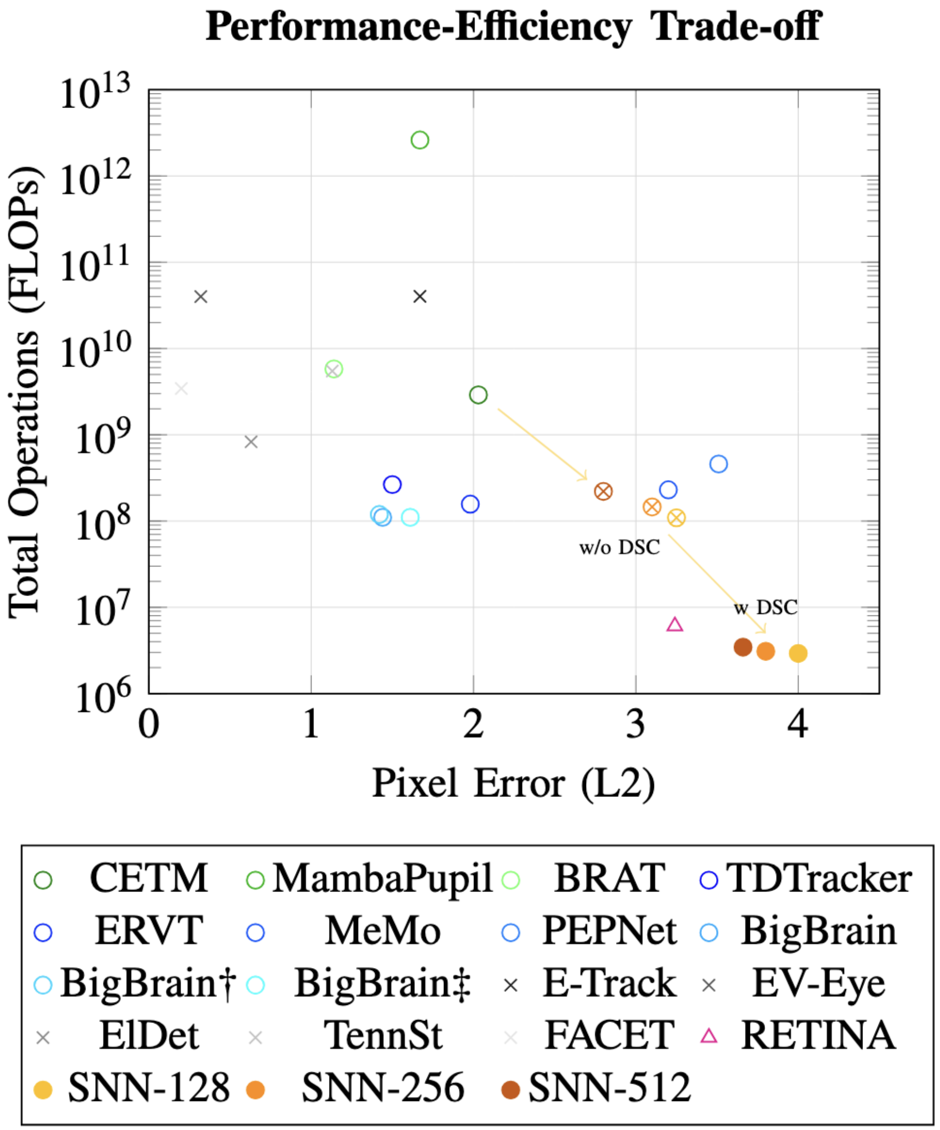
# Limitations & Future Directions




- Remaining accuracy gap to best offline ANN models, partly due to benchmark mismatch between offline metrics and real-time constraints.
- Need full hardware deployment to validate projections and uncover system-level bottlenecks.
- Future: adapt other strong architectures (BigBrain, FACET) and integrate temporal filtering akin to Retina to further stabilise predictions.
- Deploy/test in glasses-mounted environment

# Conclusions

- Projected Energy Consumption of 5mW and latency of 3 ms at 1 kHz (estimated based on SENECA neuromorphic system [Tang, 2023])
- Reduce FLOPs by 850x (2.9GFlops vs. 3.4MFlops) and parameters by 20x (7.1M vs. 353k) compared to original ANN versions
- Similar Accuracy Trade-Off as other neuromorphic eye trackers
- SNNs + neuromorphic hardware offer significant energy reductions, while sustaining a mean pixel error of 3.5-4.1px, suitable for wearable edge devices





Tang, G., et al (2023, May). *Open the box of digital neuromorphic processor: Towards effective algorithm-hardware co-design*. In *2023* IEEE ISCAS.