Towards Context-Aware Clinical Motion Understanding in Daily Living at Home: Freezing of Gait Detection with Egocentric Vision

V. Stefanova*, D. Lamsal*, M. Genbrugge , M. Yudayev , C. Schlenstedt , M. Gilat , B. Vanrumste†, and B. Filtjens†

* Equal contribution, † Equal senior contribution

ECCV 2026 | MALMÖ | SEPT 8–12

arXiv:2608.13283

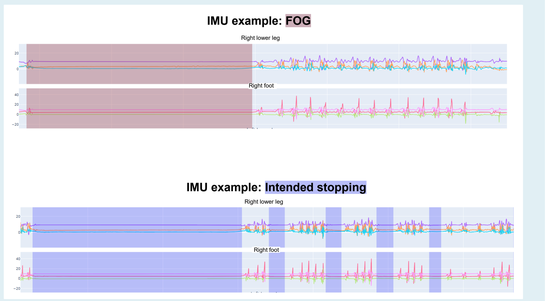
Problem & Motivation

Why this matters

- Daily-life motion is ambiguous: similar low-motion inertial signals can mean **intentional stopping, object interaction, or pathological impairment**
- Important for **real-time closed-loop assist-as-needed interventions** (e.g. cueing, exoskeletons)
- **Egocentric vision** offers contextual motion information

Case: Freezing of gait (FOG) in daily living

FOG is context-dependent: triggered by doowrays, narrow/cluttered spaces, specific home “hotspots”, etc.



Objectives

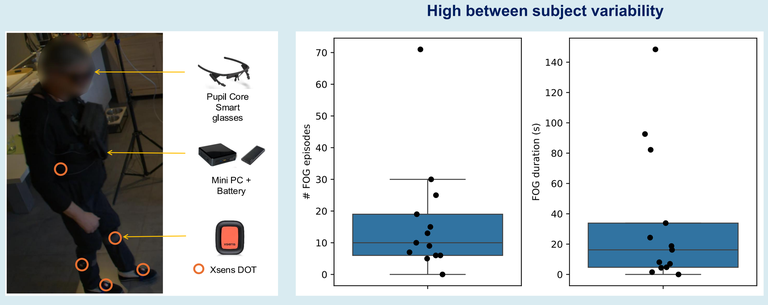
Clinical motion understanding in daily living at home: FOG detection

- First to present **egocentric vision** as a new **modality for FOG**
- Evaluate **frozen video and IMU foundation model** representations
- Qualitative analysis of **modality-specific performance**

Methods

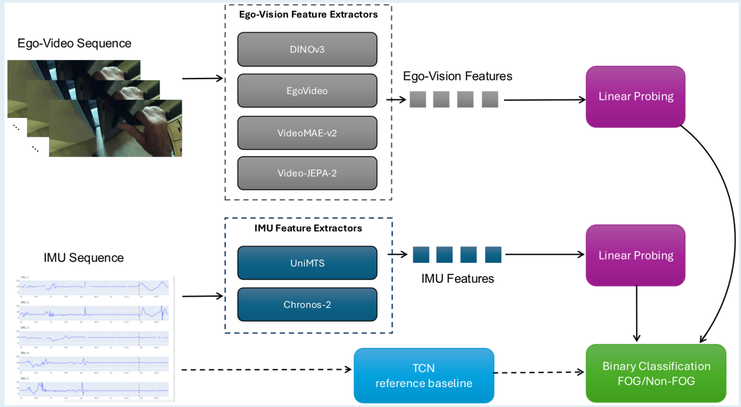
Dataset

Synchonized egocentric videos + 5 IMUs during activities of daily living at home from 15 Parkinson’s disease freezers



Experimental setup

Window-based (2s, 3s, 10s) model evaluation under leave-one-subject-out cross validation



Results

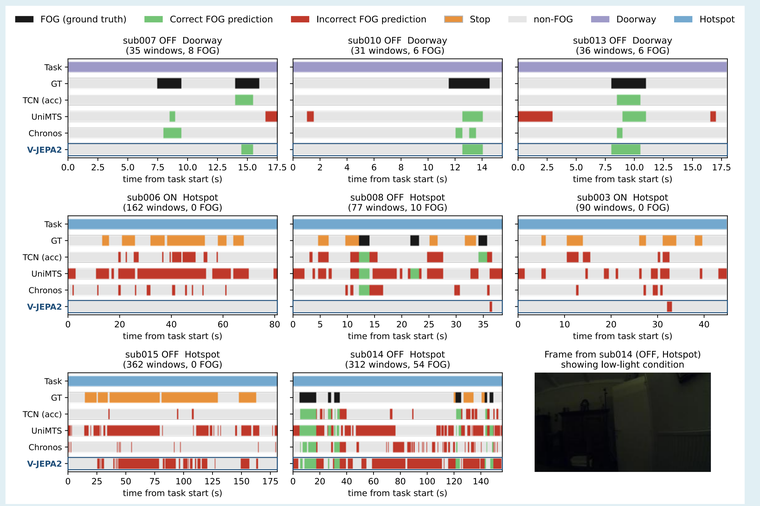
FOG detection

- IMU-based detection outperforms egocentric vision alone
- Egocentric video is predictive of FOG
- Frozen foundation model are comparative to fully trained IMU baseline

Model	F1↑	Recall↑	FPR↓	AUPRC↑	AUROC↑
Frozen foundation models - IMU					
UniMTS	29.1	**56.7**	28.8*	28.9*	73.5*
Chronos-2	**38.7**	39.9	**6.1**	**42.0***	**82.9**
Frozen foundation models - Vision					
DINOv3	17.0*	28.1	19.0*	15.4*	53.6*
VideoMAE-v2	22.0	32.8	19.3	21.8*	68.6*
EgoVideo	27.7	52.3	26.7*	23.9*	72.4*
V-JEPA2	32.6	47.4	17.6	33.0*	77.2*
Fully trained IMU baseline					
TCN (acc only)	42.3	48.6	7.2	47.3	83.0
TCN (acc+gyro)	35.5	36.6	5.8	40.0	83.3

Qualitative analysis

Egocentric video contains independent predictive information



Conclusions

- Egocentric vision from frozen foundation model representations shows potential for FOG detection at home
- Using frozen foundation model representations is comparable to training an IMU-based TCN from scratch

Future work

- IMU + Egocentric video fusion for FOG detection during daily living
- Foundation model adaptation to FOG
- Real-time closed loop implementation of an assist-as-needed intervention