# Cognitive Load Estimation through Eye-Tracking in Industrial Tasks

Kanan Gurbanov[1], Cedric Bobenrieth[1,2], Nathalie AlMakdessi[1,2], Farid Kacimi[1,2], Grégoire Chabrol[1,2], Samy Rima[3], Rabih Amhaz[1,2]

[1] ICube Laboratory, UMR 7357, University of Strasbourg, France
**amhaz@unistra.fr**
[2] Icam site de Strasbourg-Europe, France
**name.surname@icam.fr**
[3] Schmid Research Group, University of Fribourg, Switzerland
**samy.rima@unifr.ch**

**Abstract.** Understanding and managing cognitive workload is critical for safety and efficiency in industrial settings, yet practical measurement techniques are limited on the factory floor. We present a vision-based approach to estimate cognitive load using eye-tracking, hand tracking, and object detection, avoiding intrusive sensors like EEG. We evaluated this method in a real-world assembly task performed by an expert under two conditions: an organized workspace and a disorganized workspace. A mobile eye-tracker recorded the worker's gaze, while computer vision detected hands and tools in use. The disorganized condition elicited higher visual workload, evidenced by more frequent fixations and saccades, broader gaze dispersion, and more attention to irrelevant areas, despite little change in physiological proxies such as pupil size and blink rate. These results demonstrate that our non-intrusive, vision-only system can distinguish cognitive workload differences in an industrial task, laying the groundwork for in-situ workload monitoring without requiring cumbersome biosensors.

**Keywords:** Cognitive Workload, Eye-Tracking, Object Detection, Industrial Ergonomics, Human-Machine Interaction.

## 1 Introduction

Cognitive overload occurs when an operator's mental resources are exceeded by task demands, often leading to errors or slowdowns. In industrial environments, human performance under high cognitive load is a major factor in safety and productivity. Traditional measures (e.g., EEG-based monitoring) are impractical for real-time use on a production floor. There is a growing need for lightweight, real-time workload measures that can be seamlessly integrated into industrial workflows.

Many existing frameworks for workload detection rely on multimodal biosignals (EEG combined with eye-tracking or other sensors) and complex lab setups. Moreover, prior models often do not target the predominantly visual-perceptual nature of tasks like assembly and inspection. In such tasks, an operator's cognitive load may stem largely from visual search and attention management, rather than abstract reasoning alone. This work explores whether reliable indicators of cognitive workload can be obtained from vision-based data alone, specifically using an operator's eye movements, hand interactions, and observed tool usage, without the need for EEG or other invasive sensors.

We developed a proof-of-concept approach combining mobile eye-tracking with computer vision to monitor a worker's visual attention and actions during an assembly task. An expert technician

performed the same cart assembly task under organized and disorganized conditions to isolate the effect of workspace layout on cognitive load.

We hypothesized that the unorganized workspace would impose higher cognitive load, observable as more extensive visual search behavior (e.g. more fixations, wider gaze spread, longer task time), whereas the organized workspace would allow more efficient, focused visual attention.
In this paper, we report how eye-tracking metrics and interaction data differed between these conditions, and what they reveal about cognitive workload. To our knowledge, this is one of the first demonstrations of using eye-tracking with scene object detection to assess cognitive load in an in-situ industrial assembly task. The results show that subtle changes in workspace layout can be quantified through vision-based measures. Our work contributes an integrated analysis pipeline for multimodal workload monitoring and empirical evidence that a non-intrusive, video-based approach can capture cognitive load differences in real manufacturing conditions.

## 2 Related Work

This section reviews relevant literature in three domains: (1) eye movement as an indicator of cognitive load, (2) multimodal and physiological workload measurement, and (3) vision-based monitoring in industrial settings. We summarize prior achievements and identify the remaining gaps that motivate our study.

### Eye Movement as a Proxy for Cognitive Load

Eye movements (fixations, saccades, pupil dilation, microsaccades) are widely used in cognitive and HCI studies to infer mental processing (e.g. decision-making, memory retrieval), as gaze is tightly linked to attention and information acquisition. A classic study by Krejtz et al. (2018) compared pupil dilation and microsaccade metrics in arithmetic tasks, finding that both signals discriminated task difficulty, though each had distinct sensitivity profiles. [1] Many studies assume a monotonic increase in pupil diameter with load (the so-called task-evoked pupillary response). More recently, some works differentiate intrinsic vs. extraneous cognitive load using oculometric signals. For instance, a 2025 study used eye-tracking, heart-rate variability, and galvanic skin response to classify intrinsic vs. extraneous load in multimedia tasks, achieving promising predictive power. [7] This is relevant: our manipulation (workspace clutter) is akin to extraneous load, and we show gaze metrics reflect it (even when physiological signals don't shift strongly). In serious game contexts, other researchers have grounded measurement in theory, calibrating eye metrics to time-based resource-sharing models of load. E.g. a study with 42 participants playing time-critical resource-management games mapped attentional demand to gaze statistics. [6] Such work shows the feasibility of interpreting gaze within formal models of load, but still typically in constrained, quasi-lab settings rather than real-world manufacturing. Thus, while ocular metrics are well explored, there are few studies that apply them in real industrial task contexts, especially combining gaze with scene understanding (object interactions, hand movements) to localize sources of load.

### Multimodal and Physiological Workload Measurement

To capture cognitive load robustly, many works fuse multiple sensors. For example, eye tracking combined with EEG or ECG yields higher classification accuracy in cognitive load estimation than

any single modality alone. One survey notes fusion of eye, EEG, and GSR is common in ergonomics and usability studies [8]. A recent system-level work, CLERA (2023) in [2], proposes a unified deep model for joint eye-region analysis and cognitive load estimation "in the wild" (i.e. less constrained settings). CLERA jointly learns eye landmarks, blink prediction, pupil estimation, and cognitive load, and shows improved robustness over isolated pipelines. However, CLERA is built for HCI-style tasks, not industrial hand–tool workflows. It does not explicitly model object interactions or hand movement in the scene, which are crucial in an assembly task. Another approach, in the VR/training domain [9], fuses gaze with heart rate variability (HRV) and other signals to detect both cognitive load and stress, enabling adaptive systems. These systems are promising, but again not tailored for noisy, real-world factory conditions with dynamic tool usage and occlusions. Thus, while multimodal methods improve robustness, they often sacrifice deployability. Our approach intentionally remains vision-only, trading off some signal richness for practicality.

Vision-Based Monitoring in Industrial Environments

A separate but related branch of research uses computer vision to monitor worker behavior, safety compliance, and ergonomics. For instance, in construction, vision systems detect helmet usage, unsafe postures, or proximity violations [10]. A 2024 work introduces a vision-based framework for human behavior monitoring in a car door assembly line, combining multi-camera video and 3D motion capture to build a dataset (CarDA) for analyzing assembly actions. [4] That system, however, focuses on what actions are performed (pose, object presence), not why (i.e. underlying cognitive load). Other recent vision-based industrial systems target ergonomic risk assessment using skeleton estimation and motion features to flag musculoskeletal risk. For example, Agostinelli et al. (2024) propose semi-automated ergonomic risk assessment in manufacturing using depth sensors and computer vision. [11] These systems are orthogonal to cognitive load measures, they do not access gaze or attention. A recent review on computer-vision-based biomechanical and workload assessment highlights that many methods infer physical or biomechanical load (e.g. motion energy, joint torque proxies) but rarely estimate cognitive or perceptual load from video alone [3]. Furthermore, these reviews note that fusing vision with physiological or behavioral signals remains an underexplored frontier.

We can conclude from related work that most eye-tracking research remains in lab or UI settings; few studies embed gaze in complex physical tasks. Multimodal methods are powerful but often less deployable, especially outside controlled environments. Vision-only systems in industry focus on motion or safety, not cognitive load or attention dynamics. No prior work (to our knowledge) fuses gaze, hand movement, and object detection in a real industrial assembly task to estimate cognitive load, especially with controlled manipulations of environmental clutter. Our study tries to fill this gap: by combining eye-tracking with scene-level understanding (hands, tools) in a real task, we demonstrate that gaze metrics can reliably detect visual workload changes in situ, bridging the sensor-rich lab world and the vision-only industrial world.

## 3 Method

Participants & Task. We recruited one expert assembly technician, Robert Bolusset (LEAN manufacturing consultant and trainee), to perform a cart assembly task under two workspace conditions: organized (tools neatly laid out) and unorganized (same tools, scattered randomly) cf.

Fig 1. The task procedure remained identical in both trials, ensuring any behavioral differences stem from layout effects.

**Fig. 1.** Example frames from organized vs unorganized videos

**Apparatus & Data Collection.** The participant wore a mobile Pupil Labs eye tracker, capturing synchronized scene video and gaze overlay streams (fixations derived from raw gaze). We processed video frames with MediaPipe for hand detection and YOLOv8 (trained on 850+ annotated images across 11 object classes) for tool/part detection (achieving ~87% mAP at the validation threshold). We synchronized gaze, fixation ID, object presence, and hand interaction per frame into a structured JSON timeline dataset.


**Fig. 2.** Our expert during the assembly process wearing pupil lab neon Eye-tracker.

**Measures & Analysis.** From this timeline, we extracted:

- **Eye metrics:** total fixations, average fixation duration, saccade count, gaze dispersion, pupil diameter, blink rate
- **Attention allocation:** proportion of fixations on relevant vs. irrelevant areas, AOI (area of interest) transition frequency, dwell times on key objects
- **Task performance:** total completion time, any corrections or tool mis-selections

We compared these metrics between organized and unorganized trials in a within-subject, descriptive analysis.

## 4 Results

The disorganized workspace produced clear, consistent divergences in gaze behavior and attention distribution relative to the organized layout:

- **Fixation & saccade counts** increased by ~24%, from 494 to 613 fixations, and 493 to 612 saccades, respectively, indicative of a more fragmented scanning strategy.
- **Average fixation duration** dropped slightly from ~563 ms (organized) to ~550 ms (unorganized), reflecting shorter glimpses and more frequent transitions.
- **Gaze dispersion** expanded from a radius of ~158 px to ~190 px, confirming broader spatial search.
- **Attention misallocation:** The share of fixations landing on irrelevant areas rose from 62.0% to 73.6%, confirming that clutter forces more "wasted" visual glances.
- **AOI transition frequency** also rose, signaling more gaze switching between items and distractors.
- **Pupil diameter** and **blink rate** remained effectively unchanged (~5.23 mm, ~11 blinks/min), suggesting that while visual scanning load increased, overall arousal or stress did not escalate detectably for this expert.
- **Completion time** increased in the disorganized condition (fitting with the increased visual workload), though no severe errors or tool misuses occurred, consistent with the participant's expertise.

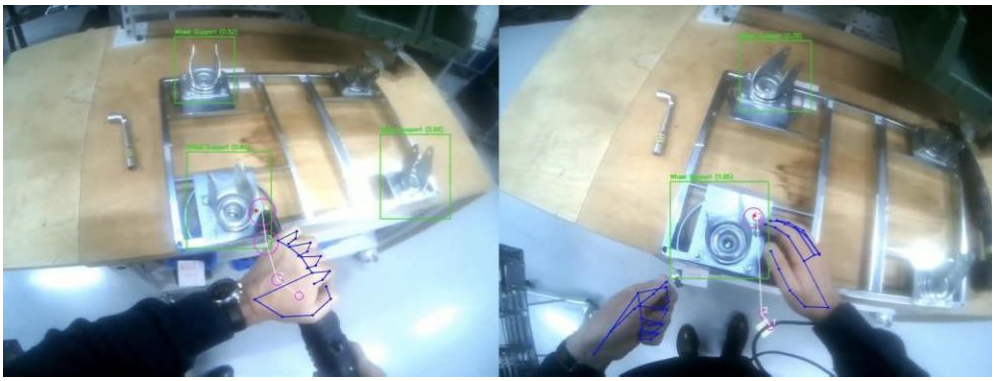

**Fig. 3.** Result video combining data between industrial object detection, hand tracking, and eye fixation and transition of the eyes (purple circles, size of the circles is the duration of the fixation)

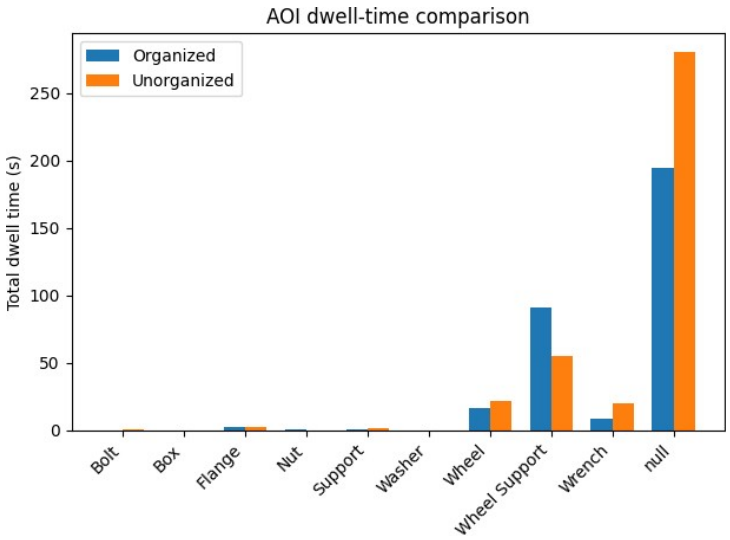


**Fig. 4.** AOI dwell-time comparison between organized and unorganized workplace correlated with industrial object detection

These differences align with theoretical expectations: increasing extraneous visual demands (via clutter) forces more scanning and attentional shifts, which we were able to detect through gaze metrics alone. The absence of strong changes in pupil/blink measures implies that for an expert operator, the extra visual load remained manageable in terms of overall cognitive stress.

## 5 Discussion

Our results validate that workspace layout directly influences visual/perceptual load, observable through eye metrics. In cognitive load terms, the disorganized layout introduced extraneous load, effort invested in navigating the environment rather than intrinsic task logic. The increase in fixations, saccades, gaze dispersion, and AOI switching signals this overhead.

Interestingly, the stable physiological metrics (pupil, blink) suggest that the mental load ceiling was not breached. Because our participant was an expert, they likely compensated for extra visual demands without triggering elevated arousal. This aligns with research showing that experts often absorb increased perceptual demands via compensatory strategies before cognitive overload manifests (e.g. in pupil dilation) in novices.

Relative to prior work, Krejtz et al. demonstrated that pupil and microsaccade signals track mental difficulty, but they do so in controlled tasks, not physical tasks with environmental clutter. CLERA offers an elegant vision-only model for cognitive load in “the wild,” but lacks a coupling to object interaction or hands, something essential in assembly tasks. Works in industrial vision largely focus on action or safety detection rather than the why (i.e. internal cognitive strain). For example, CarDA monitors worker actions but not attention or cognitive load.

Thus, our work treads a new space: vision-enabled cognitive load estimation in a manual industrial context. We show that eye–scene fusion yields signals of perceptual workload in a realistic environment.

Limitations:

- Single expert subject: Results reflect conditions for a highly skilled user; novices might show different coupling between visual and physiological signals.
- Task specificity: We tested one assembly scenario (cart). Generalizing to other tasks (e.g., inspection, wiring) requires further validation.
- Lighting and tracking noise: Real workshop conditions can degrade gaze or object detection; occasional frame drops or misalignments may introduce noise.

Implications & Insights

Deployment feasibility: Our approach (eye + video) is far less intrusive than EEG, making it a more viable candidate for industrial deployment.
Explainability of load signals: Since we link gaze to objects and hands, we can interpret where the extra load originates (e.g. scanning distractors), not just infer a black-box workload score.
Potential for adaptive feedback: Even without live feedback in this study, the pipeline could support continuous monitoring and alerts when visual load becomes excessive.

## 6 Conclusion

We have demonstrated that a vision-based system, combining mobile eye-tracking, hand detection, and object recognition, can detect meaningful differences in visual workload arising from workspace organization in an industrial task. Despite the same underlying assembly procedure, the cluttered layout imposed higher perceptual demand, evident via gaze metrics (fixation count, dispersion, AOI distribution). Crucially, this was achieved without physiological sensors.

Our findings suggest that vision-only workload sensing is viable in pragmatic settings, especially for tasks where visual search and attention dominate. The next steps include validating across more participants and tasks, calibrating thresholds to individual operators, and exploring real-time deployment.

## References


1. Krejtz, K., Duchowski, A. T., Niedzielska, A., Biele, C., & Krejtz, I. “Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze.” PLoS ONE, 2018. Link: https://doi.org/10.1371/journal.pone.0203629
2. Ding, L., Terwilliger, J., Parab, A., Wang, M., Fridman, L., Mehler, B., & Reimer, B. (2023). CLERA: A unified model for joint cognitive load and eye region analysis in the wild. ACM Transactions on Computer-Human Interaction. https://doi.org/10.1145/3603622
3. Egeonu, D., & Jia, B. (2025). A systematic literature review of computer vision-based biomechanical models for physical workload estimation. Ergonomics, 68(2), 139–162. https://doi.org/10.1080/00140139.2024.2308705

4. Papoutsakis, K., Bakalos, N., Fragkoulis, K., Zacharia, A., Kapetadimitri, G., & Pateraki, M. (2024). A vision-based framework for human behavior understanding in industrial assembly lines. arXiv preprint arXiv:2409.17356. https://arxiv.org/abs/2409.17356
5. Martinez-Cedillo, A. P., Gavrila, N., Mishra, A., Geangu, E., & Foulsham, T. (2025). Cognitive load affects gaze dynamics during real-world tasks. Experimental Brain Research, 243(4), 82. https://doi.org/10.1007/s00221-025-07037-4
6. Sevcenko, N., Appel, T., Ninaus, M. et al. Theory-based approach for assessing cognitive load during time-critical resource-managing human–computer interactions: an eye-tracking study. J Multimodal User Interfaces 17, 1–19 (2023). https://doi.org/10.1007/s12193-022-00398-y
7. Ekin, M., Krejtz, K., Duarte, C. et al. Prediction of intrinsic and extraneous cognitive load with oculometric and biometric indicators. Sci Rep 15, 5213 (2025). https://doi.org/10.1038/s41598-025-89336-y
8. Skaramagkas, V., Giannakakis, G., Ktistakis, E., Manousos, D., Karatzanis, I., Tachos, N. S., Tripoliti, E., Marias, K., Fotiadis, D. I., & Tsiknakis, M. (2023). Review of eye tracking metrics involved in emotional and cognitive processes. *IEEE Reviews in Biomedical Engineering, 16*, 260–277. https://doi.org/10.1109/RBME.2021.3066072
9. Nasri, M. (2025). Towards Intelligent VR Training: A Physiological Adaptation Framework for Cognitive Load and Stress Detection. arXiv preprint arXiv:2504.06461. https://doi.org/10.48550/arXiv.2504.06461
10. Cheng, J. C. P., Wong, P. K.-Y., Luo, H., Wang, M., & Leung, P. H. (2022). Vision-based monitoring of site safety compliance based on worker re-identification and personal protective equipment classification. *Automation in Construction, 139*, 104312. https://doi.org/10.1016/j.autcon.2022.104312
11. Agostinelli, T., Generosi, A., Ceccacci, S. et al. Validation of computer vision-based ergonomic risk assessment tools for real manufacturing environments. Sci Rep 14, 27785 (2024). https://doi.org/10.1038/s41598-024-79373-4