Abstract

Eye movement analysis plays a critical role in understanding human visual attention, perception, and cognition. The accurate detection of fundamental eye movement events: fixations, saccades, smooth pursuits (SPs), and post saccadic oscillations (PSOs) is essential for reliable interpretation of eye-tracking data across fields such as psychology, neuroscience, and human computer interaction. Event detection, however, remains challenging due to the variability of gaze patterns across tasks, the inherent difficulty of distinguishing short duration events such as PSOs, and the lack of standardized evaluation protocols. This thesis addresses these challenges by conducting a comprehensive investigation of event detection methods, spanning threshold-based, machine learning, and deep learning models, with a particular focus on generalizability across datasets, feature selection, and the methodological importance of PSO detection.

The work begins by systematically evaluating a wide range of event detection algorithms using high-resolution eye tracking data. Traditional threshold-based methods such as IVT and IDT were benchmarked against machine learning approaches (Random Forest) and deep learning models (CNN and LSTM). Unlike prior comparative studies that often used different datasets or evaluation criteria, all algorithms here were evaluated under identical conditions using manually annotated ground truth labels and consistent performance metrics. Results confirmed that threshold-based methods are efficient for binary classification of fixations and saccades but are highly sensitive to parameter tuning and unable to generalize well. In contrast, machine learning and deep learning methods achieved superior robustness and higher agreement with human coders, with CNNs showing the best overall accuracy. These findings establish a clear methodological baseline for the advantages of data-driven models while also demonstrating the critical role of evaluation methodology particularly cross-validation strategies—in producing reliable results.

Building on this foundation, the thesis proposes a hybrid 2D-CNN-LSTM network for simultaneous classification of fixations, saccades, PSOs, and SPs. The CNN layers act as spatial feature extractors, while the LSTM layers capture temporal dependencies, making the architecture well-suited for sequential gaze data. To assess the impact of input representation, four kinematic features: velocity, acceleration, jerk, and direction were systematically combined into feature sets (VD, VAD, VJD, VAJD). The analysis revealed that classification performance is strongly feature-dependent: combinations including velo-

city and direction with either acceleration or jerk produced the most robust results, while the full four-feature set did not consistently improve accuracy. Importantly, this feature analysis shed light on the persistent challenge of PSO detection. Although PSOs were often confused with short saccades or fixations, feature sets incorporating jerk improved detection rates, reaching 67% accuracy compared to substantially higher performance for fixations, saccades, and SPs. These findings underscore the difficulty of PSO classification but also demonstrate the promise of hybrid models in approaching this challenge.

The thesis next examines the cross-task generalization of eye movement event detection. Eye movement data were analyzed across diverse visual tasks, including reading, static image viewing, video watching, and moving-dot tracking. Statistical analyses confirmed that oculomotor behavior is task-dependent: reading is characterized by rapid fixations and saccades, dynamic stimuli elicit smooth pursuits, and post-saccadic oscillations (PSOs) show task-specific variability. A CNN trained and tested across tasks demonstrated high within-task accuracy but suffered significant degradation in cross-task transfer, especially for PSOs. These results highlight the limitations of current models when applied beyond their training domain and point to the necessity of developing domain-generalized event detection approaches capable of handling heterogeneous visual and cognitive contexts.

Finally, the thesis turns to the applied significance of PSO detection, focusing on reading research. Using CNN-based classification, results were compared against commercial software that ignores PSOs and labels only fixations and saccades. The findings showed that excluding PSOs inflates fixation durations and alters widely used reading metrics such as average fixation length, which are critical for evaluating text complexity and cognitive processing. By incorporating PSOs into detection, the analysis produced more accurate and nuanced measures of reading behavior, confirming that PSOs are not merely artifacts but meaningful components of gaze dynamics. This provides an important methodological insight: ignoring PSOs risks systematic misinterpretation of eye-tracking studies in reading and related domains.

In summary, this thesis contributes: (i) a rigorous benchmarking of classical, machine learning, and deep learning algorithms under standardized conditions; (ii) the introduction of a hybrid CNN-LSTM framework with systematic feature analysis; (iii) empirical evidence for the critical importance of PSO detection in applied reading studies; and (iv) one of the first comprehensive evaluations of cross-task generalization in event detection. Collectively, the findings advance the methodological foundations of eye movement research and have broad implications for building adaptive, accurate, and context-aware eye movement analysis systems for diverse scientific and applied domains.