Modeling Eye Gaze Trajectories with Spectrally Regularized GANs

Shailendra Bhandari[1], Pedro Lencastre[1], Alex Szorkovszky[2], Anis Yazidi[1], Pedro Lind[1,2]
[1]Oslo Metropolitan University, [2]SimulaMet

OSLOMET
OSLO METROPOLITAN UNIVERSITY
STORBYUNIVERSITETET

Introduction

Eye gaze trajectories provide critical insights into human visual attention, perception, and cognitive processes [1, 2]. Modeling the dynamics of stochastic processes, such as eye gaze trajectories, is a significant challenge with practical applications in fields like human-computer interaction, cognitive science, and artificial intelligence (AI). These trajectories are essential for understanding behaviors in tasks like reading, visual search, and interaction with computer systems [3]. The sequential nature and inherent stochasticity of eye movement patterns pose significant challenges for accurate modeling and prediction. Traditional models, such as Markov models, offer approximate estimations but often fall short in accounting for the complex interdependencies influenced by perceptual constraints, memory, and cognitive factors [4]. Recent advancements in AI and deep learning have significantly propelled the field of generative models [5, 6]. Among these methods, Generative Adversarial Networks (GANs) have achieved remarkable success [7, 8].

Objective: Develop models that accurately capture the sequential nature of eye gaze velocities during a visual search task by representing these velocities as stochastic processes.

Methods

Generative Adversarial Networks (GANs)

GANs comprise two neural networks—the generator (G) and the discriminator (D)—trained concurrently. The generator attempts to create fake samples to deceive the discriminator, which learns to distinguish between real and generated samples.

To enhance the training of the generator, we complement its loss function with spectral loss L_{spectral} [9, 10]. We also applied Markov models and Hidden Markov Models (HMMs) to model the eye gaze velocity data, capturing the temporal dependencies and stochastic properties inherent in eye movements.

References

[1] Kowler, E. (2011). Eye movements: The past 25 years.
[2] Duchowski, A. T. (2002). A survey of eye-tracking applications.
[3] Shapiro, S. C. (2001).
[4] Patterson, T. A. et al. (2008). State–space models of movement.
[5] Harshavardhana, B. R. et al. (2020). Survey on generative models.
[6] Oussidi, M., & Elhassouny, A. (2018). Deep generative models.
[7] Goodfellow, I. et al. (2014). Generative adversarial nets. *NeurIPS*, 2672-2680.
[8] Gui, J. et al. (2020). GANs: Review and applications.
[9] Mathieu, M. et al. (2016). Deep video prediction.
[10] Durall, R. et al. (2020). Watch your up-convolution.
[11] SR Research. (2024).
[12] Handford, M. (1987). *Where's Waldo?*

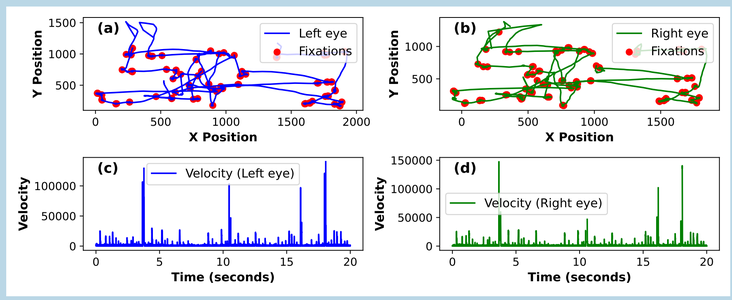
Acknowledgments

This work is funded by the Research Council of Norway under grant number 335940 for the project *Virtual-Eye*.

Data and Statistical Measures

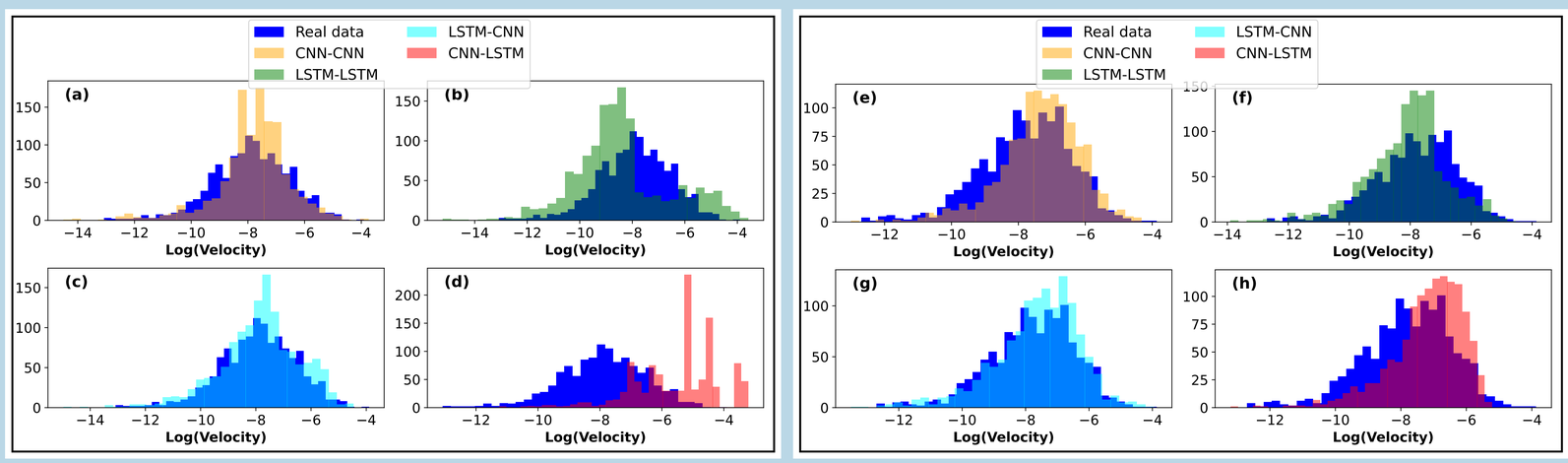
Data Collection

Eye-tracking data were collected at Oslo Metropolitan University using the Eye-Link Duo device [11]. Participants performed a visual search task with images from *Where's Waldo?* [12]. We used the Jensen-Shannon divergence (D_{JS}) and spectral loss (L_{spectral}) to evaluate model performance. To quantify how well the time dependencies are replicated, we evaluated the autocorrelation of the velocity magnitude $\|\mathbf{v}\|$.



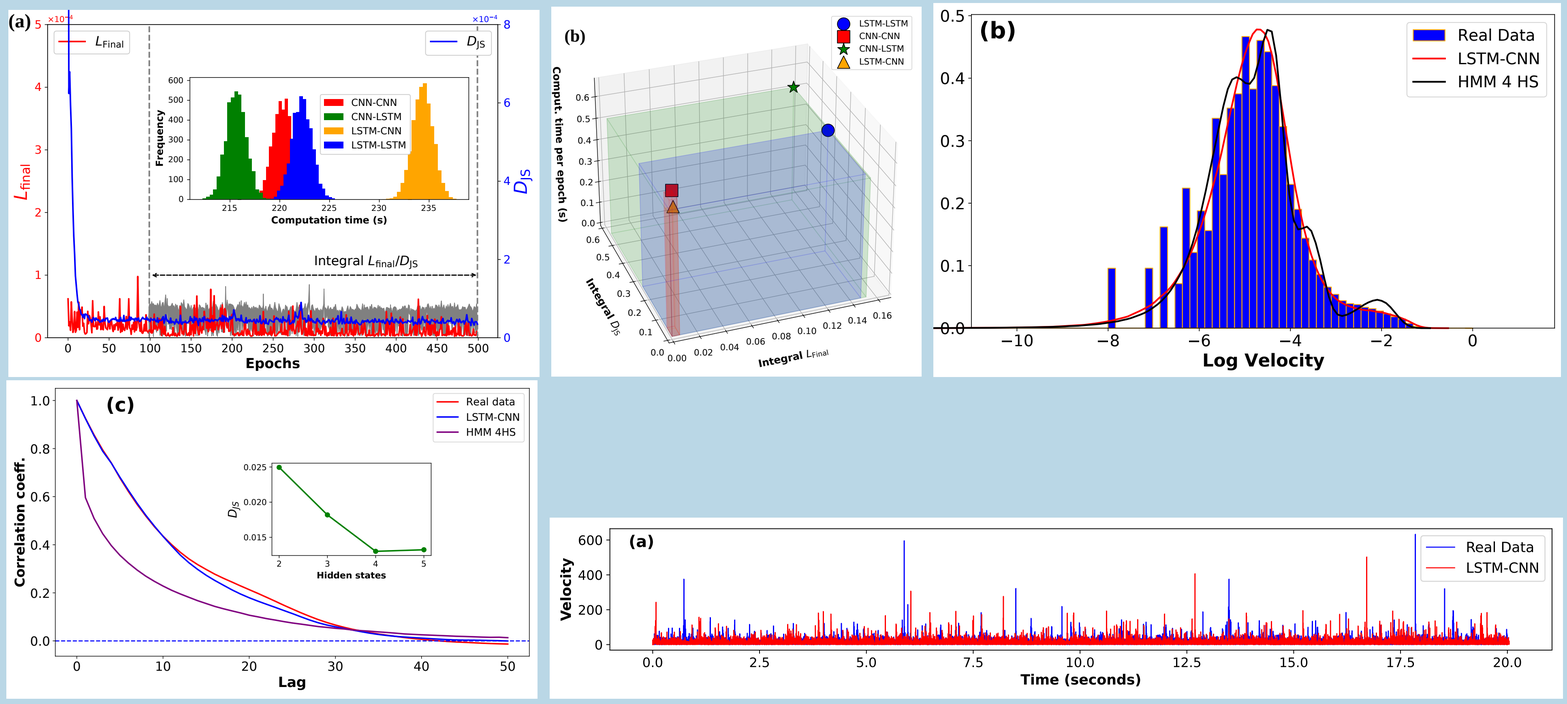
Results

Architecture Selection Experiment



Among the four GAN architectures, the LSTM-CNN model with spectral loss (L_{final}) showed the closest alignment to the real data distribution, especially in the distribution tails.

Model Performance Comparison



The LSTM-CNN GAN model significantly outperformed the Hidden Markov Model (HMM) in capturing the complex temporal and distributional characteristics of eye gaze velocity data.

Conclusion

Our study demonstrates that spectrally regularized LSTM-CNN GANs effectively model eye gaze velocity trajectories, outperforming traditional Markov models. The LSTM-CNN GAN achieved the lowest Jensen-Shannon divergence ($D_{JS} = 0.00032$), indicating exceptional alignment with the real data distribution.

Key Insights: **Architectural Synergy**: Combining LSTM layers in the generator with CNN layers in the discriminator leverages the strengths of both architectures.

Spectral Loss: Emphasizes the frequency components of the data, ensuring accurate modeling of both high-frequency and low-frequency elements.

Computational Efficiency: Maintains computational efficiency despite increased complexity, balancing training overhead with performance gains.

These findings have significant implications for developing simulation environments, training systems, and eye-tracking technologies that rely on realistic synthetic data.