

DEGREE PROJECT

L

Synthetic Mouse Trajectory Generation Using GANs

Maximillian Pavliska

Computer Science and Engineering, bachelor's level
2025

Luleå University of Technology
Department of Computer Science, Electrical and Space Engineering



[This page intentionally left blank]

ABSTRACT

Behavioral biometrics use mouse dynamics as a method for user authentication and bot detection. Real mouse movement data collection at scale proves difficult because of privacy issues, time requirements and diversity limitations.

This research investigates the application of Generative Adversarial Networks (GANs) to produce artificial mouse movement trajectories that mimic human like behavior. The WGAN-GP model used LSTM networks to generate synthetic mouse movement data from sequential movement information. The model received training from the Bogazici Mouse Dynamics dataset before undergoing evaluation through statistical metrics, classifier performance assessments and visual inspection.

The generated data matches real samples in both spatial and temporal characteristics. The generated data proved difficult for shallow classifiers to differentiate but deep models like LSTMs demonstrated the ability to detect differences through their analysis of timing patterns.

The research shows GANs can effectively produce realistic behavioral data while showing that generating human like motion in sequential models remains an ongoing challenge.

CONTENTS

CHAPTER 1 – THESIS INTRODUCTION	1
1.1 Introduction	1
1.2 Motivation	2
1.3 Problem Definition	2
1.4 Delimitations	3
1.5 Thesis Structure	4
CHAPTER 2 – BACKGROUND	7
2.1 Behavioral Biometrics	7
2.2 Mouse Dynamics	7
2.3 Generative Adversarial Networks (GANs)	8
2.4 Mouse Movement Datasets	9
2.5 Machine Learning and Evaluation of Mouse Trajectories	10
2.6 Challenges in Mouse Movement Generation	11
CHAPTER 3 – RELATED WORK	13
3.1 Behavioral Biometrics and Mouse Dynamics	13
3.2 Challenges in Data Collection and Synthetic Generation	14
3.3 Generating Synthetic Mouse Trajectories	14
3.4 Positioning the Current Work	15
CHAPTER 4 – METHOD	17
4.1 Dataset and Preprocessing	17
4.2 Model Architecture: GAN for Mouse Trajectory Generation	18
4.3 Evaluation Metrics	18
CHAPTER 5 – RESULT	21
5.1 Overview	21
5.2 Quantitative Evaluation	21
5.3 Visual Evaluation	23
CHAPTER 6 – DISCUSSION	31
6.1 Spatial Realism	31
6.2 SVM Discriminability: Mid-Level Realism	31
6.3 Deep Model Detection	31
6.4 Answers to Research Questions	32
6.5 Sustainability and Ethical Considerations	34

6.6 Summary and Insights	35
CHAPTER 7 – LIMITATIONS	37
7.1 Training Time and Resource Constraints	37
7.2 Limited Hyperparameter Exploration	37
7.3 Mode Collapse and Instability	37
7.4 Evaluation Scope	38
7.5 Dataset Limitations	38
7.6 Theoretical vs. Practical Performance Gap	38
CHAPTER 8 – FUTURE WORK	39
8.1 Model Architecture Improvements	39
8.2 Exploring TimeGAN for Temporal Realism	39
8.3 Training Optimization and Stability	40
8.4 Adversarial Feedback with Stronger Discriminators	40
8.5 Expanded Evaluation Protocols	40
8.6 Dataset Diversity and Generalization	40
8.7 Ethical Considerations and Responsible Use	41
CHAPTER 9 – CONCLUSION	43
REFERENCES	45

CHAPTER 1

Thesis Introduction

1.1 Introduction

As human-computer interaction becomes increasingly central to web and software experiences, it has become essential to understand user behavior for cybersecurity, human-centered design and behavioral analytics applications. The study of mouse movement patterns through mouse dynamics serves as a fundamental behavioral biometric which identifies distinct user and group behaviors [8].

The collection of mouse movement data serves multiple applications including user authentication systems, fraud detection systems and behavioral modeling systems [3]. The process of obtaining high-quality extensive datasets containing real user mouse trajectories faces multiple obstacles [1]. The process of data collection extends over time while providing limited behavioral variety and raises privacy and ethical issues. Such data limitations hinder the development of reliable models that depend on this information.

The use of Generative Adversarial Networks (GANs) provides an effective solution through synthetic data generation that replicates real-world distributions [4]. The application of GANs to mouse trajectory generation allows for the production of realistic and diverse user interaction data that protects user privacy [10]. The synthetic data serves multiple purposes including dataset enhancement, detection system testing and human behavior simulation in controlled settings.

This research investigates the application of Generative Adversarial Networks to produce artificial mouse movement trajectories which mimic real human input behavior. The research develops a generative model that reproduces human mouse movement patterns both spatially and temporally before assessing the generated data realism through statistical analysis, visual inspection and machine learning tests. This research develops sustainable ethical solutions for behavioral data generation which benefit research and

practical applications.

This work is intended for developers, researchers and students working in synthetic data generation and behavioral biometrics. In particular those working with machine learning models for input behavior

1.2 Motivation

The growing sophistication of automated bots makes traditional detection methods such as CAPTCHAs, IP filtering and browser fingerprinting less effective [6][5]. The detection method of behavioral biometrics through mouse dynamics presents an effective solution. The systems analyze human-computer interaction patterns through timing and curvature and speed variability which bots find challenging to duplicate[3].

The development of behavior-based detection or authentication models requires extensive access to diverse datasets containing real user interactions. The process of obtaining this data requires extensive time and produces limited results while raising privacy and ethical issues since it involves recording user input across web platforms.

Generative Adversarial Networks (GANs) present a potential answer through their ability to generate synthetic data that mimics human behavior. The successful implementation of these models would decrease the need for invasive data collection while allowing extensive training of detection algorithms and enabling controlled assessments of system resistance to synthetic inputs[6]. The research aims to enhance behavioral dataset accessibility while studying how machine learning models generate artificial human input.

1.3 Problem Definition

In the field of behavioral biometrics, mouse movement analysis plays a crucial role in distinguishing between human users and automated agents. The development of systems for user authentication, behavior prediction and bot detection requires high-quality mouse movement datasets. The process of collecting such data at scale faces major obstacles due to it being time consuming and shows limited behavioral variety while creating privacy and ethical issues.

The technology of Generative Adversarial Networks (GANs) demonstrates its ability to generate synthetic data that mimics intricate patterns. The application of GANs to mouse trajectory generation would overcome real-world data collection constraints by allowing the production of extensive customizable datasets. The generated datasets serve multiple purposes including training detection systems, simulating user input and security research applications.

This thesis addresses the problem of generating synthetic mouse movement trajectories that closely mimic human behavior. The research focuses on answering these three research question:

1. How can a GAN-based model be designed to generate realistic and behaviorally diverse mouse trajectories in terms of spatial and temporal characteristics?
The research aims to generate data that maintains both realistic characteristics and diverse patterns.
2. Do the generated trajectories match real human mouse movement data and to what degree, both statistically and visually?
The research evaluates the similarity between synthetic data and real user input through statistical and model-based methods.
3. Can the generated trajectories be used as an effective machine learning tool for classification and anomaly detection purposes. Either by replacing or enhancing real data?
The research examines the suitability of this data for machine learning applications.

1.4 Delimitations

The research investigates the generation of mouse movement data that mimics human movement through Generative Adversarial Networks (GANs). The project established multiple boundaries which determined its scope and maintained control over its size based on time and resource availability.

- **Single Dataset:** All experiments were conducted using the Bogazici mouse dynamics dataset. The research results lack generalization because it used only the Bogazici mouse dynamics dataset without incorporating any other mouse movement data sources.
- **Fixed-Length Sequences:** The evaluation of mouse movement data consisted of 50 event sequences with fixed lengths. The evaluation process did not include variable-length or real-time sequences because it required training and evaluation simplification.
- **Offline Evaluation:** The evaluation of generated data was performed using offline techniques such as classifier accuracy, statistical distance metrics, and visualization. Real-time or live user testing was not included.
- **Limited Hyperparameter Search:** A complete hyperparameter optimization process was not possible because of time limitations. The research focused on investigating WGAN-GP with LSTM layers as the single GAN architecture.

- **Training Constraints:** GAN training requires substantial computational resources and extended periods of time for completion. The limited computational resources allowed only a few complete training runs to be finished. The need to restart training when mode collapse occurred restricted the number of experiments that could be conducted.
- **Evaluation Scope:** The evaluation of generated data included realism assessment and classifier performance evaluation but excluded application-based assessments such as training downstream models with synthetic data.

1.5 Thesis Structure

The thesis consists of nine chapters which work together to study the application of Generative Adversarial Networks (GANs) for generating synthetic mouse trajectories that replicate human movements.

- **Chapter 1 – Introduction:** Presents the research subject while explaining the reasons behind the study, defines the problem and establishes the limitations.
- **Chapter 2 – Background:** Explains foundational concepts related to behavioral biometrics, mouse dynamics, Generative Adversarial Networks (GANs) and essential machine learning methods.
- **Chapter 3 – Related Work:** Reviews previous research about synthetic data generation and mouse movement analysis to establish its position among existing studies.
- **Chapter 4 – Method:** Describes the dataset, preprocessing steps, model architecture, and the evaluation metrics for assessing data realism.
- **Chapter 5 – Results:** Presents the outcomes of quantitative and qualitative evaluations, including classifier accuracy, statistical distances, and visualization of generated versus real data.
- **Chapter 6 – Discussion:** Interprets the results in terms of spatial and temporal realism, classifier performance, and overall model effectiveness.
- **Chapter 7 – Limitations:** Discusses the constraints encountered during the project, such as training instability, limited experimentation time, and dataset limitations.
- **Chapter 8 – Future Work:** Proposes directions for expanding the research, including architectural improvements, use of alternative models like TimeGAN, stronger evaluation frameworks, and broader applications.

- **Chapter 9 – Conclusion:** Summarizes the findings and contributions of the thesis, reflecting on the feasibility and potential of GAN-based synthetic mouse trajectory generation.

CHAPTER 2

Background

The purpose of this section is to explain the concepts that are investigated in this thesis about distinguishing between human mouse movements and GAN-generated synthetic movements. This section introduces behavioral biometrics through mouse dynamics before explaining GAN data generation principles and provides a brief overview of machine learning methods for classification purposes.

2.1 Behavioral Biometrics

Behavioral biometrics refers to the measurement and analysis of human characteristics for identification or authentication purposes[14]. Behavioral biometrics tracks patterns that emerge from human actions while physiological biometrics depend on physical traits such as fingerprints and iris patterns[14][8]. The patterns which behavioral biometrics measure develop through time and include gait patterns and voice patterns as well as keystroke dynamics and mouse dynamics[8].

Behavioral biometrics provide multiple valuable benefits within the cybersecurity space. The unobtrusive nature of behavioral patterns captured through mouse and keyboard input result in a cost effective and user friendly system compared to sensor based systems [8][12]. It enables continuous authentication across the entire session duration instead of only during login[8]. However, behavioral patterns show greater variability between different sessions of the same user compared to physiological characteristics which creates difficulties for recognition systems [8].

2.2 Mouse Dynamics

The study of mouse dynamics involves analyzing the behavioral patterns which users demonstrate when they operate computer systems through mouse or trackpad pointing devices. The analysis includes both specific mouse actions such as clicks (single, double,

right clicks) and scrolling as well as the essential aspects of cursor movement between screen points [8].

A mouse movement trajectory exists as a sequence of (x,y) coordinates which get recorded at timestamp (t). The raw data allows researchers to derive multiple features which describe movement characteristics [8].

These features can be described as follows.

- **Kinematic Features:** The physical aspects of movement can be described through velocity, acceleration, jerk (acceleration change rate), curvature, direction and movement angle [8][1][5].
- **Statistical Features:** The calculation of mean, standard deviation, minimum, maximum, skewness and kurtosis values from kinematic features extends across entire trajectories and their specific segments [8][1].
- **Geometric Features:** The trajectory shape can be described through five key features which include total distance traveled, straight line distance between endpoints, straightness index (ratio of straight line to traveled distance), number of pauses, and deviation from a straight path [8][5].

2.3 Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs) is a powerful class of deep learning models designed for generative modeling and consists of two independent neural networks that train simultaneously. The two neural networks are the Generator (G) and the Discriminator (D) [4][7].

- **Generator (G):** The Generator networks primary task is understanding real data distribution patterns and produces synthetic samples that mimic those patterns. The Generator takes a random noise vector as input and attempts to transform it into a synthetic sample that resembles the real data.
- **Discriminator (D):** The Discriminator network works as a binary classifier in its design. The Discriminator exists to distinguish real samples from the synthetic samples produced by the Generator.

The training process resembles a zero-sum game. The Generator works to produce increasingly more realistic data and tries to deceive the Discriminator while the Discriminator works on trying to improve its ability to detect synthetic data from the Generator. In this training loop both networks improve their performance simultaneously [4][7].

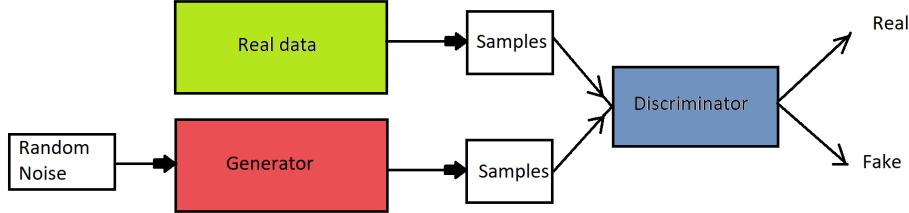


Figure 2.1: A GAN model simplified

2.4 Mouse Movement Datasets

Mouse movement datasets are crucial for training and evaluating both detection and generation models involving mouse dynamics. Both the quality and quantity of datasets essentially define how good a model will be. The Bogazici mouse dynamics dataset[9] is used in this study for its detailed time-stamped xy coordinates alongside button states and active window data from user browsing sessions.

BeCAPTCHA-Mouse[1] introduced an open-source dataset which includes both real human and GAN-generated mouse movements to help researchers study detection systems under adversarial conditions.

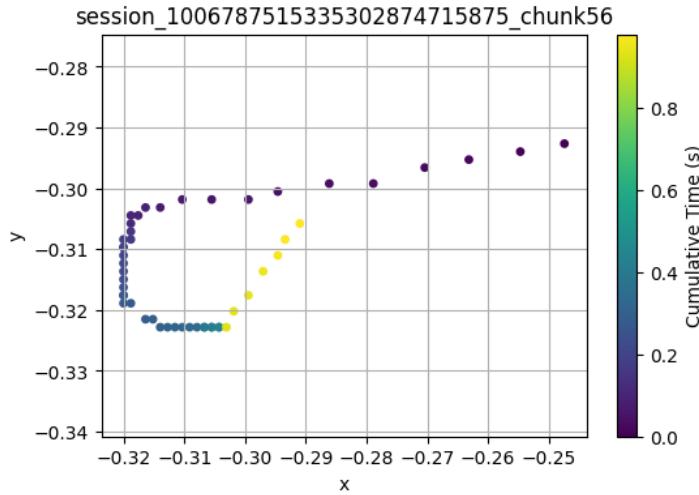


Figure 2.2: A visualization of a mouse trajectory

2.5 Machine Learning and Evaluation of Mouse Trajectories

While the main objective of this thesis is to generate realistic mouse movement data using a GAN, machine learning methods remain essential for evaluating the quality and realism of the synthetic trajectories. The evaluation process includes statistical comparisons together with classification-based models that learn to identify real versus fake data.

- **Traditional Classifiers with Feature Engineering:** As discussed in Section 2.2, several features can be extracted from mouse trajectories. These features can then be fed into standard classifiers such as SVM, Random Forest, or Gradient Boosting models [8][3].
- **Recurrent Neural Networks (RNNs):** Long Short Term Memory (LSTMs) and Gated recurrent units (GRUs) are specifically designed to handle sequential data. They possess internal memory mechanisms that allow them to learn long-range temporal dependencies directly from the raw sequence and often bypassing the need for manual feature extraction[11][6].
- **Convolutional Neural Networks (CNNs):** While primarily known for image processing, CNNs can be used in 2 different approaches. The first approach is plotting the sequential data of mouse trajectories into 2D images and processing the images for classification[11][6]. The second approach is using 1D-CNNs, this approach can be effective for sequences by learning local patterns or motifs over time. They can be used alone or in a hybrid architecture together with RNNs [1][11].

The choice of model involves trade-offs between the need for feature engineering, the ability to capture complex temporal patterns, and computational requirements.

2.6 Challenges in Mouse Movement Generation

The creation of realistic synthetic mouse trajectories involves several technical and ethical challenges:

- **Data Diversity:** The wide range of human behaviors creates difficulties in developing models that avoid overfitting problems [6][8][11].
- **Ethical and Privacy Concerns:** The continuous behavioral monitoring of users raises privacy issues which require both transparent policies and user consent agreements[6][8].

CHAPTER 3

Related Work

The research field of human-computer interaction, behavioral biometrics and machine learning now depends on realistic human behavior simulations. Mouse movement data contains extensive behavioral information which reveals subconscious decision-making processes and physical motor patterns and intentional actions. The initial research on mouse dynamics concentrated on bot detection and authentication but current studies investigate data generation methods to address real-world data collection constraints[1][8].

This chapter examines the existing literature about behavioral biometrics and mouse dynamics modeling and GANs as tools for creating synthetic mouse movement data that replicates human behavior patterns.

3.1 Behavioral Biometrics and Mouse Dynamics

Behavioral biometrics refers to how users interact with the application, especially the dynamics of the computer mouse and keyboard. This data can be used to recognize patterns and distinguish bots from real human users. Human movements often contain minor speed fluctuations together with directional and acceleration patterns that bots struggle to mimic naturally[1][11].

Early research [2] proved mouse dynamics as a valid behavioral biometric system, by using movements such as movement speed, direction, and distance to create user profiles. Subsequent research expanded the feature sets by adding more complex kinematic features such as velocity and acceleration as well as statistical measures like mean, variance, skewness and kurtosis extracted from trajectory segments[8][1][5].

Research by [1] and their development of machine learning models built with mouse trajectory features (velocity, acceleration, and angle changes) demonstrated a more than 99% success rate in detecting bots from a single move movement.

The work of [11] showed that the visual patterns found in mouse trajectories when plotted on a graph can also be used for bot identification by using convolutional neural networks (CNNs) for image based recognition.

3.2 Challenges in Data Collection and Synthetic Generation

The process of acquiring high-quality mouse trajectory datasets for research faces multiple obstacles. The collection of data faces three main obstacles which include privacy issues, ethical problems, and the requirement of extensive manual data acquisition. The Bogazici Mouse Dynamics Dataset [9] serves as a useful real-world sample but its scalability and generalizability remains limited for various use cases.

Researchers have developed synthetic generation methods to address these limitations. Traditional methods for trajectory generation used rule-based or functionally defined trajectories which included linear, sinusoidal and exponential paths together with handcrafted velocity profiles [13]. The traditional methods failed to reproduce the complex unpredictable and varied human input behaviors that occur naturally.

To tackle this challenge, the focus has shifted towards synthetic data creation as an answer to the requirement for more realistic data and to extend the limited human datasets. The BeCAPTCHA-Mouse dataset [1] presented an approach that included both functional generation through linear, quadratic, and exponential functions with multiple different velocity profiles as well as GAN-based data generation for mouse trajectory synthesis. The study revealed that GAN-generated data samples combined with functional synthetic data samples significantly improved both bot detector training and performance [1].

3.3 Generating Synthetic Mouse Trajectories

Obtaining high-quality labeled data for both human and bot interaction is a well-known challenge in the field. While datasets such as the Bogazici Mouse Dynamics Dataset [9] offer real user interactions, the acquisition of realistic bot movement data has historically relied on rule-based or simplistic trajectory generation methods [13].

To address this, recent research has turned to synthetic data generation, both for improving training and for stress-testing detection systems. [1] introduced a hybrid framework that combined functionally generated movements (linear, exponential, etc.) with adversarially generated samples using GANs. This dataset, known as BeCAPTCHA-Mouse, was used to evaluate bot detection systems under adversarial conditions [1].

The foundational work of [4] introduced Generative Adversarial Networks (GANs) as a means of learning data distributions through adversarial training between a Generator

and a Discriminator. GANs have since been applied in multiple domains for generating realistic behavioral data.

[6] used a Deep Convolutional GAN (DCGAN) to generate image-like representations of mouse movements. Their results showed that these GAN-generated samples drastically reduced detection performance, from 99.9% down to 45.2% recall on previously robust CNN-based detection systems. This demonstrates the potential for generative models to imitate human behavioral signals convincingly enough to fool state-of-the-art classifiers.

3.4 Positioning the Current Work

Previous research has shown that GANs can be used to synthesize mouse behavior either as static images or trajectory features, but most studies have emphasized evasion and focusing on generating data to defeat classifiers. This thesis, by contrast, focuses on the generation of temporally realistic, continuous mouse trajectories in raw (x, y, deltaT) format using a WGAN-GP architecture with recurrent networks.

While [1] and [6] demonstrated the vulnerability of classifiers to GAN-based samples, their models did not produce time-series outputs directly usable for simulation or sequence modeling. This work builds on those foundations by designing a sequence-based GAN model that captures both spatial and temporal characteristics of human behavior. It further evaluates the generated data using multiple realism indicators. This includes classifier confusion, statistical distribution alignment, and visual inspections, rather than detection accuracy alone.

The goal is not to bypass security systems, but to better understand the limits of behavioral biometrics and to create realistic synthetic data for future robustness testing, training, and simulation research.

CHAPTER 4

Method

4.1 Dataset and Preprocessing

The research utilizes the Bogazici mouse dynamics dataset [9] to analyze human mouse movement recordings. The dataset consists of sessions from different users, each session represents a sequence of mouse events with the following attributes (client_timestamp, x, y, button, state, window) as data points.

- The preprocessing of this dataset removed the button, state and window attributes as this was not needed for the research of generating mouse trajectories.
- Each session was divided into chunks (sections) containing 50 events as fixed segments.
- The client_timestamp was converted to deltaT (time difference between event[A] and event[A-1]) and all chunks containing a deltaT value greater than one second were removed for ensuring consistency throughout the study.
- The coordinates (x, y) received normalization processing which transformed them to the [-1, 1] range based on the monitor resolution (1920x1080).

After preprocessing the data, the total amount of chunks was approximately 18.000 real human trajectories. The GAN model was then trained on 10.000 chunks and the remaining 8000 were used for evaluation purposes. The preprocessed trajectory data exists as files in .pt format for both training and evaluation purposes with dimensions [50, 3].

4.2 Model Architecture: GAN for Mouse Trajectory Generation

The Generative Adversarial Network (GAN) was implemented to generate realistic mouse trajectories. The architecture of the model follows the Wasserstein GAN with Gradient Penalty (WGAN-GP). This model was chosen due to its ability to generate realistic data with improved convergence properties. The model also is more stable during training with gradient penalty and more robust against issues like mode collapse.

- **Generator:** The generator uses an LSTM as its network architecture to process 100-dimensional noise vectors. The output sequences are of length 50 steps while each step generates x, y coordinates together with deltaT values.
- **Discriminator (Critic):** The Discriminator functions as a binary classifier that uses an LSTM structure to determine sequence authenticity. The training method uses Wasserstein loss with gradient penalty to enforce the Lipschitz constraint.

The training process extended for 10,000 epochs with the Adam optimizer ($lr=1e-4$, $\text{betas}=(0.5, 0.9)$), the critic also received more updates than the generator ($n_{\text{critic}} = 3$).

4.2.1 Training Environment

The GAN was implemented in Python using the PyTorch framework. All experiments in this thesis were conducted using the following hardware and software setup:

- **GPU:** NVIDIA GeForce RTX 3070 with 8 GB VRAM
- **CPU:** AMD Ryzen 7 5800X (8 cores / 16 threads @ 3.80 GHz)
- **RAM:** 32 GB DDR4
- **Operating System:** Windows 11
- **Frameworks:** Python 3.10.11, PyTorch 2.5.1, CUDA 12.1

Each version of the GAN model was trained for between 3,000-10,000 epochs (based on mode collapse). The training of a single model required between 16-48 hours depending on both the batch size and network architecture. The hardware limitations together with the long training duration of GANs restricted the number of parameter settings that could be tested within the project schedule.

4.3 Evaluation Metrics

Both quantitative and qualitative assessment methods were used to evaluate the realism of the generated data.

4.3.1 SVM Discriminability Test

A linear Support Vector Machine (SVM) received training to recognize whether generated trajectories came from real data. A 50% accuracy rate from the SVM classifier signifies that the generated data has similar properties to the real data. This classifier was re-trained for each 200 GAN epoch and was trained on a balanced data set containing real human data (not the same as the GAN received for training) and generated samples from each epoch. SVM was selected as a simple classical linear classifier to evaluate the generated data.

4.3.2 Wasserstein Distance

The 1D Wasserstein distance calculated the feature-wise statistical similarity between real and generated data across x, y and deltaT dimensions. The measurement decreases when distributional alignment becomes more effective. This type of measurement is well suited to measure distribution overlap in generative models.

Baseline values for Wasserstein distance

- Real vs Real = 0,001 - 0,01
- Real vs Random noise = 0,5 - 1,0

4.3.3 t-SNE and UMAP Visualization

The real and generated trajectories underwent 2D projection through t-SNE (t-distributed Stochastic Neighbor Embedding) and UMAP (Uniform Manifold Approximation and Projection). The presence of real and fake data point clusters which overlap in the visualization indicates equivalent latent structure between them. This offers an intuitive sense of whether real and generated data samples cluster similarly.

4.3.4 Deep Classifier Test

A small LSTM classifier received training to identify sequences from real data and generated data. A deep model's performance near 50% accuracy indicates that it matches complex temporal patterns successfully. This test learns realism at a higher level of complexity and is able to identify more subtle timing and movement patterns.

4.3.5 Qualitative Visualization

The analysis included both real and generated trajectory visualizations for inspecting their curvature patterns and timing behaviors and overall naturalness. Human evaluators checked the sequences for their natural appearance.

4.3.6 Alternative Evaluation Approaches

Methods such as Fréchet Distance, DTW (Dynamic Time Warping), Perceptual User Studies, or another machine learning classifier could have further improved the evaluation of the generated data. However, due to scope and time limitations this study focused on well-established techniques.

CHAPTER 5

Result

5.1 Overview

The evaluation of the generative model for producing human-like mouse trajectories is presented in this chapter. The realism of the generated data was evaluated through both quantitative metrics (classifier confusion, statistical similarity) and qualitative analyses (visual inspection, t-SNE projections). The results provide insight into the effectiveness of the WGAN-GP architecture and its capacity to replicate the temporal and spatial characteristics of human mouse movement.

5.2 Quantitative Evaluation

5.2.1 SVM Discriminability Test

A linear SVM was trained on a balanced dataset of real and generated mouse trajectories using statistical and kinematic features. The classifier achieved an accuracy of **51.9%** at Epoch 3700, suggesting that it was only marginally better than random guessing. This near-random performance indicates that the synthetic samples exhibit similar statistical properties to real trajectories, making them difficult to separate via simple decision boundaries.

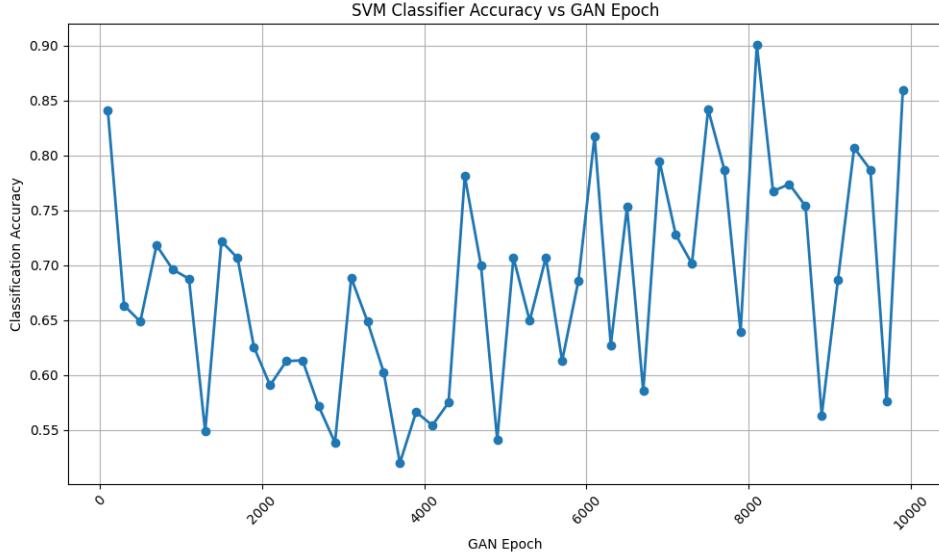


Figure 5.1: SVM Classifier Accuracy vs GAN Epochs

5.2.2 Wasserstein Distance

The 1D Wasserstein distance was used to evaluate distributional similarity between real and generated data across the **x**, **y**, and **deltaT** dimensions:

- **x**: 0.0201
- **y**: 0.0097
- **deltaT**: 0.0083

The low values demonstrate that the distributions of real and generated data match closely across all three axes. The generated mouse movements show close similarity to real positional data because the Wasserstein distances for spatial features (**x** and **y**) remain low.

The Wasserstein distance results confirm that the GAN model successfully learned to duplicate spatial aspects of human-like mouse trajectories.

5.2.3 Deep Classifier Test

To evaluate the progression of realism in the generated trajectories during GAN training, a pre-trained LSTM classifier was used to distinguish between real and generated samples at various GAN training epochs.

Figure 5.2 shows the classifier accuracy when tested against GAN outputs from epochs 100 to 9900. The LSTM model was trained in advance on a balanced dataset of real and generated sequences and remained fixed throughout this evaluation.

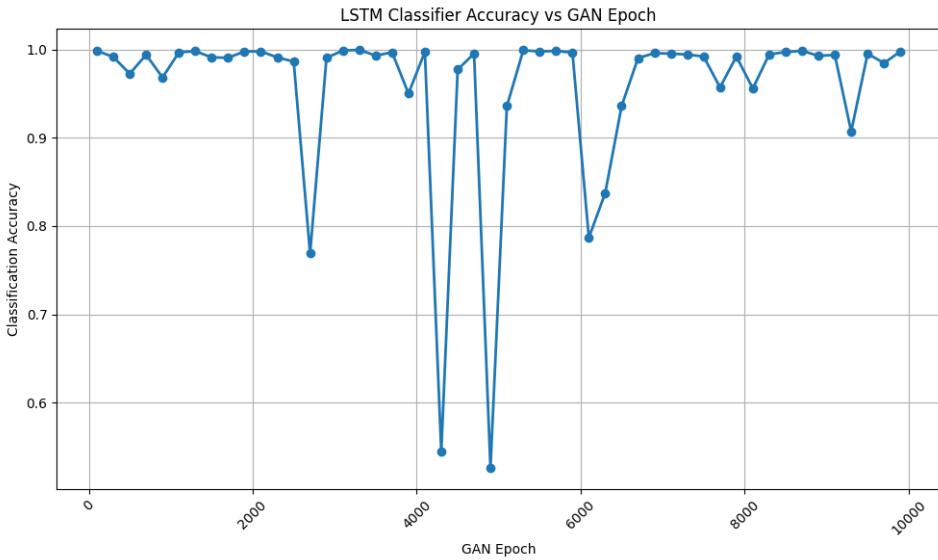


Figure 5.2: LSTM Classifier Accuracy vs GAN Epochs

- The accuracy of the classifier remained high throughout the training period of the GAN often exceeding 99%, showing that while the GAN learned to produce more complex data, the LSTM could still detect synthetic patterns reliably.
- Notable drops occurred at a few epochs (e.g., **54.45%** at epoch 4300 and **52.6%** at epoch 4900), indicating moments where the GAN briefly produced highly realistic samples that confused the classifier. However, these drops were not sustained, and the classifier quickly regained high accuracy, suggesting instability rather than a true breakthrough in realism.

These results show that although the GAN improves during training, it does not consistently fool a deep sequence model. This supports the conclusion that the temporal dynamics of mouse movements remain the most challenging aspect for the generator to replicate.

5.3 Visual Evaluation

5.3.1 t-SNE Visualization

Figure 5.3 and Figure 5.4 shows the t-SNE projection of 1.000 and 10.000 samples of real and generated sequences. A significant overlap between the two distributions was observed, supporting the quantitative results and suggesting that the generated data shares similar latent structure with real data.

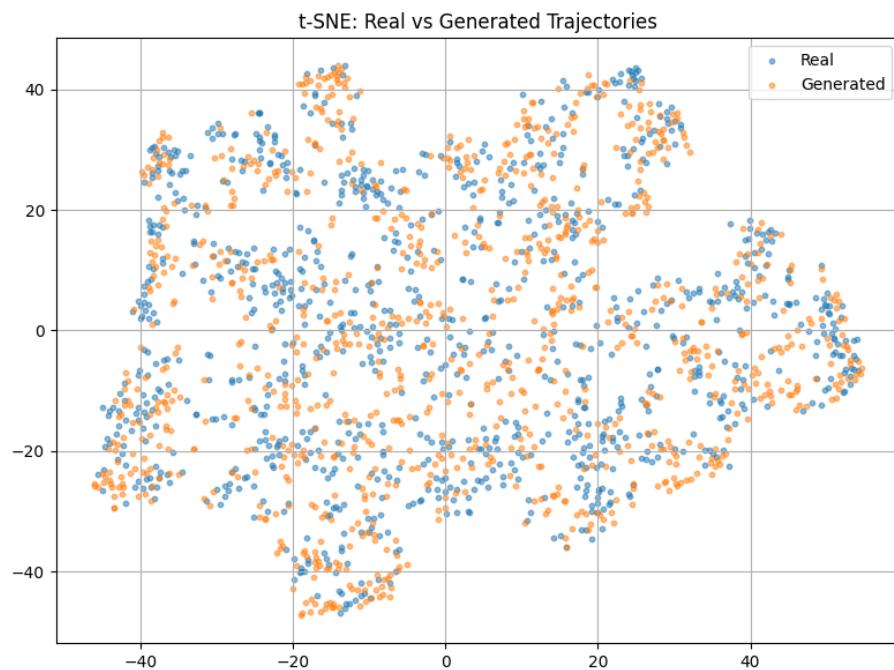


Figure 5.3: t-SNE projection of real (blue) and generated (orange) mouse trajectories (1k samples).

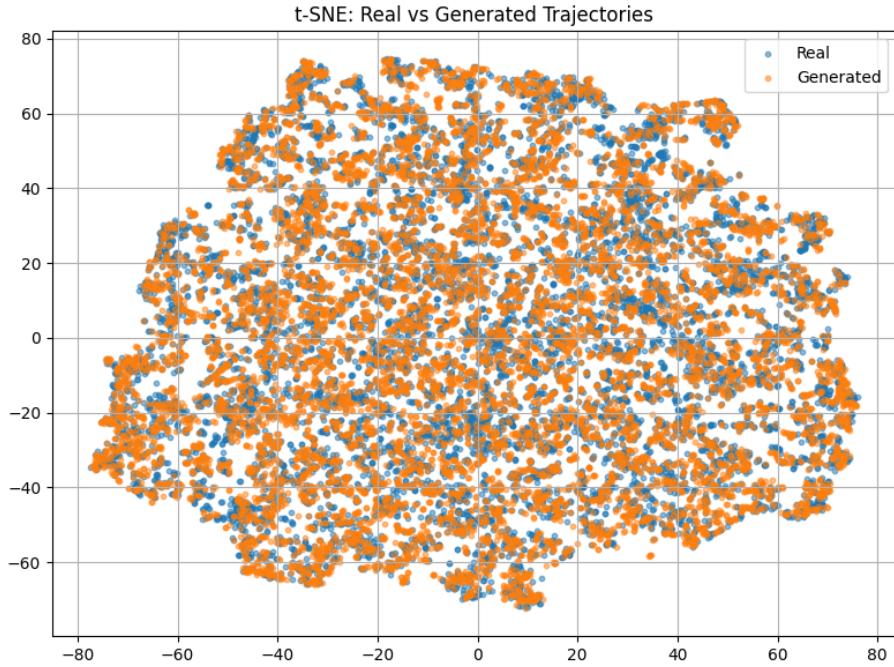


Figure 5.4: t-SNE projection of real (blue) and generated (orange) mouse trajectories (10k samples).

5.3.2 UMAP Visualization

The trajectory data also received visualization in 2D space through the implementation of Uniform Manifold Approximation and Projection (UMAP) in addition to t-SNE. The nonlinear dimensionality reduction method UMAP preserves both local and global data structures in contrast to t-SNE. This enables the detection of structural overlaps or anomalies at scale.

The UMAP projections of real and GAN-generated data appear in Figures 5.5 and 5.6 with 1.000 and 10.000 samples.

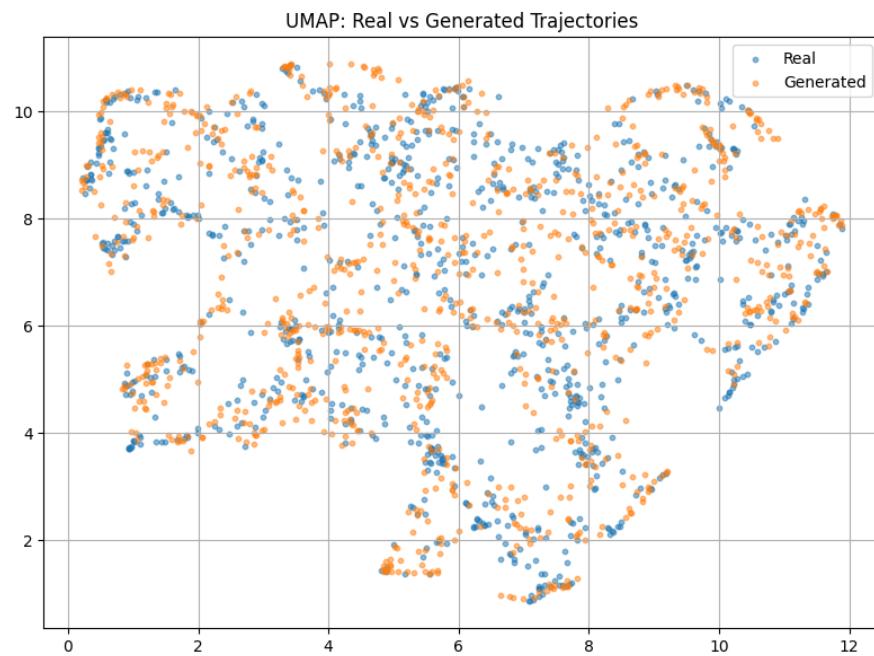


Figure 5.5: UMAP projection 1,000 samples. Some separation remains visible between real and generated data.

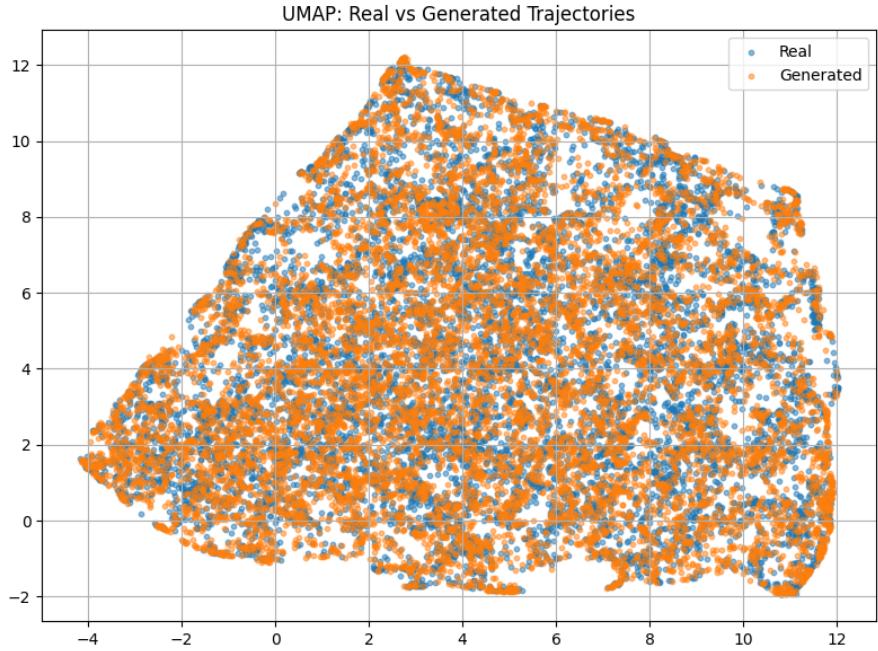


Figure 5.6: UMAP projection 10,000 samples. Real and generated samples are highly intermixed.

The clusters from real and generated samples show good alignment at the 1.000 scale but some separation exists which indicates minor distribution differences. The two datasets show complete overlap at the 10.000 scale indicating that the generator achieves good generalization and synthetic data matches the structural diversity of real human trajectories.

5.3.3 Trajectory Examples

The visual realism of generated mouse trajectories became assessable through the comparison of plotted sample sequences against actual mouse movements extracted from the dataset.

The real mouse trajectory in Figure 5.7 demonstrates natural human movement characteristics through its smooth curves and intermittent pauses and its variable speed patterns.

The GAN-generated trajectory in Figure 5.8 demonstrates paths that maintain human-like characteristics through their non-linear shape and moderate random variations.

The generated paths maintain a comparable spatial pattern to real data through their curved paths and their moderate level of randomness which helps create human-like movement patterns. The synthetic paths maintain a visual similarity to actual dataset paths through their complex structures.

The examples confirm the numerical results by demonstrating the GAN successfully learned both detailed and broad patterns found in human mouse movement data.

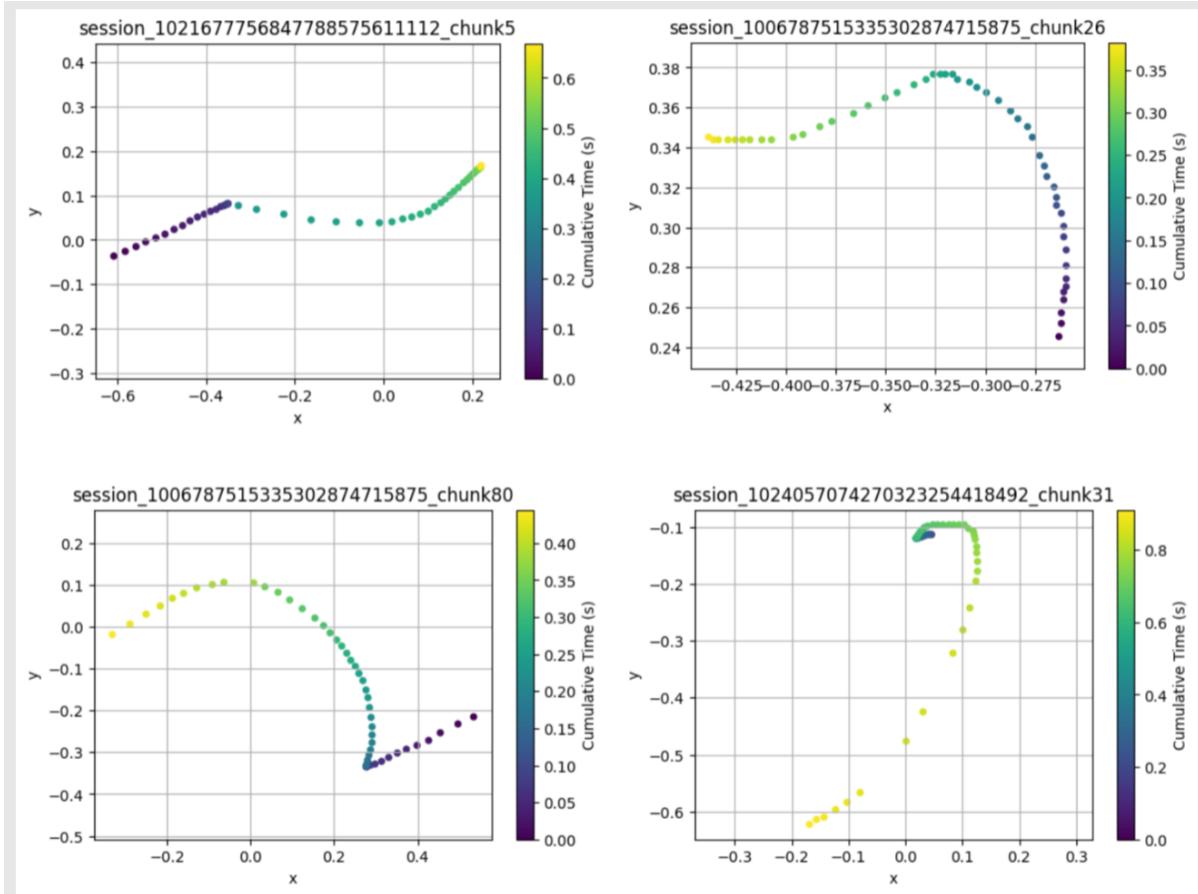


Figure 5.7: Human trajectories.

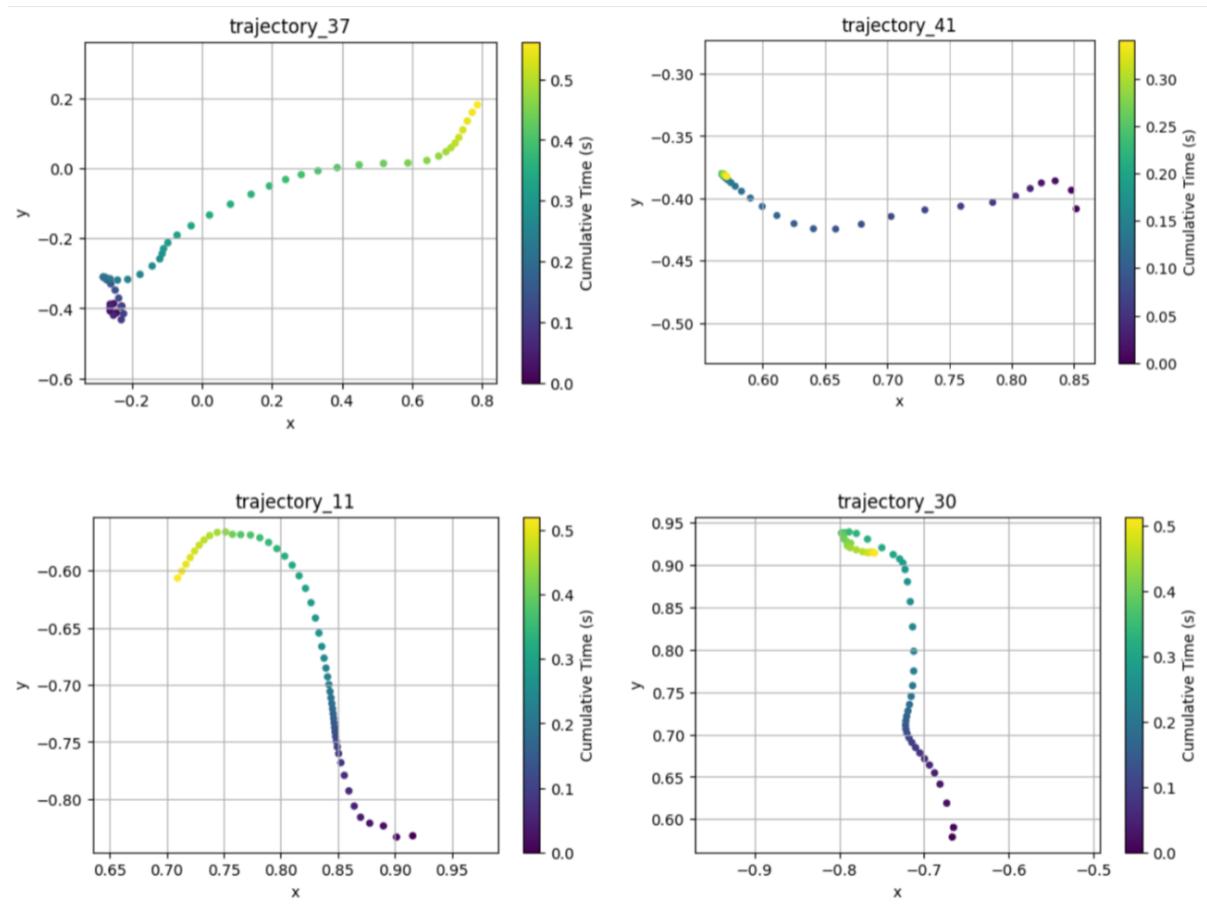


Figure 5.8: Generated trajectories.

CHAPTER 6

Discussion

This section combines and discusses the findings from the quantitative and qualitative evaluation of the generated mouse trajectories. The results show the strengths and the limitations of the GAN model and its ability to produce realistic human-like behavior.

6.1 Spatial Realism

The Wasserstein distance results show that the real and generated data are statistically very similar. The distances of 0.0201 (x), 0.0097 (y), and 0.0083 (ΔT) indicate strong alignment across spatial dimensions. These values indicate that the GAN has successfully learned the distributional properties of real user trajectories.

6.2 SVM Discriminability: Mid-Level Realism

The SVM classifier accuracy showed a continuous decline from its initial value of 84.1% at epoch 100 until it reached levels near random guessing at epochs 2900–3900. The linear model’s ability to distinguish real data from generated data decreased throughout GAN training. The classifier accuracy showed a rebound in later epochs (e.g., 81.7% at epoch 6100 and 90.0% at epoch 8100) which indicates inconsistent generative quality and possible mode collapse or overfitting behavior during longer training periods. The mid-training epochs (e.g., 2900–3900) appear to yield the most indistinguishable outputs.

6.3 Deep Model Detection

The LSTM-based deep classifier demonstrated consistent high accuracy above 99% throughout almost all training epochs. The deep model trained on temporal features demonstrates its ability to detect small timing and movement sequencing differences between real and generated data even though statistical and shallow classifiers fail to distinguish between them. The classification accuracy experienced brief drops to approximately

54.5% during GAN epoch 4300 but since these instances did not persist it suggests some training instability rather than a breakthrough. Overall, temporal coherence remains the most difficult feature for the GAN to replicate convincingly.

6.3.1 Latent Structure Analysis: t-SNE and UMAP

The GAN-generated samples' ability to duplicate real mouse trajectory structural patterns was evaluated through t-SNE and UMAP dimensionality reduction techniques.

The high sample size revealed that generated and real data points created overlapping clusters through t-SNE because of their strong feature space alignment. UMAP further supported these findings. Real and generated points demonstrated good but imperfect mixing at 1,000 samples while maintaining some distinct groupings. The distributions became almost entirely interleaved when the sample size reached 10,000. The generated samples demonstrate two capabilities: they replicate human data characteristics at the individual level and they duplicate the complete population structure of human data when the dataset size grows.

The visualizations show that the GAN successfully reproduces both local movement patterns and global behavioral diversity.

6.3.2 Qualitative Realism and Human Perception

Although not formally scored, visual inspection of the generated trajectories showed smooth, natural-looking curves, consistent direction changes, and plausible timing behavior. The visual characteristics of the generated patterns match human behavioral traits which indicates the model produces realistic patterns that appear authentic to human observers when viewed without additional context.

6.4 Answers to Research Questions

6.4.1 Research Question 1

How can a GAN-based model be designed to generate realistic and behaviorally diverse mouse trajectories in terms of spatial and temporal characteristics?

To generate realistic mouse trajectories, a WGAN-GP architecture was implemented with LSTM-based generator and discriminator networks. The selected structure maintains sequential data features and stable training operations. The model was trained on real human trajectory data which was split into segments of 50 steps. The generator learned to produce synthetic sequences with realistic spatial features and timing characteristics.

Conclusion:

The selected GAN architecture learned to generate the spatial patterns of human mouse movement. However, the process of learning behavioral diversity together with realistic temporal dynamics remains challenging. The generated data maintained authentic surface-level patterns yet sequence-based models were still able to detect the hidden temporal patterns.

6.4.2 Research Question 2

Do the generated trajectories match real human mouse movement data and to what degree, both statistically and visually?

To evaluate this multiple techniques were applied:

- The Wasserstein distance values across all dimensions were low indicating a close distributional alignment.
- t-SNE and UMAP projections showed an increasing overlap between the real and generated data as the training progressed.
- The SVM classifier accuracy dropped to 50-60%, indicating that a linear classifier struggles to distinguish differences.

Conclusion:

The generated mouse trajectories showed a strong alignment with the statistical distributions and the visual structures of the real data. The low wasserstein distance and the overlapping clusters in t-SNE and UMAP projections shows that the GAN was able to mimic the global characteristics of the real human movement.

However, experimenting with a deep LSTM model showed that the temporal aspect and behavioral complexities were not fully mimicked. This suggests that while GAN models may be able to learn patterns at surface level, deeper patterns still remain a challenge. Architectural improvements in the model may lead to more accurate data generation.

6.4.3 Research Question 3

Can the generated trajectories be used as an effective machine learning tool for classification and anomaly detection purposes. Either by replacing or enhancing real data?

A deep LSTM classifier was used to evaluate the ability to distinguish between real and synthetic data. The LSTM achieved high accuracy rates above 95%, showing it was able to recognize differences between the real and generated data.

Conclusion:

While the synthetic data shows a potential for basic models, it fails to match the real data in advanced machine learning applications. The synthetic mouse trajectories need

additional development in temporal modeling or architecture design to replace/enhance real data in classification and anomaly detection systems.

6.5 Sustainability and Ethical Considerations

While the primary focus of this thesis is about the technical aspects of using GANs for mouse movement generation, it's also important to consider the impact this has on the environment, society and privacy.

6.5.1 Technical and Economic Sustainability

The major benefit of generating synthetic data is the reduced need for collecting large amounts of real human data. This saves both time, money and could improve the training of machine learning models since labeled data is limited. This could in turn help research and development become more accessible and cost effective.

6.5.2 Environmental Impact

Training GANs takes a lot of time and computational power which also means higher electricity use. Even though this project was limited to around 6 weeks of experimentation and the amount of training was not very large. Future work using GANs should be aware of the environmental impact it has when training deep learning models and look at ways of making the process more efficient. This could be done by, for example, training models on servers that are powered by renewable energy.

6.5.3 Social and Ethical Considerations

This type of technology can be used in both good and bad ways. The positive aspect is that it can help improve research in areas such as human and computer interaction and bot detection systems. On the other hand, this technology can be misused by bad actors to generate fake human behavior and bypass security systems. Future work in this field should focus on using this type of data generation in safe and ethical ways.

6.5.4 Privacy and Individual Impact

Another benefit of generating data is that it protects users privacy, since collecting real mouse movement data from people can lead to privacy concerns. Using synthetic data instead offers a way to work with behavioral patterns without using real users information. This makes it a good option for privacy friendly research

6.6 Summary and Insights

- The GAN model achieved a strong distributional similarity (Wasserstein distance) and a mid-level fooling ability against linear classifiers (SVM)
- The effectiveness of LSTMs at detecting differences remains high because temporal irregularities continue to exist at a subtle level.
- The t-SNE and UMAP projections demonstrate promising alignment of latent structure.
- The GAN faces difficulties with training stability because realism becomes unstable during the later epochs.

The research shows that spatial realism has been nearly achieved but the temporal aspect presents an ongoing challenge for generating human-like mouse trajectory data. The research shows that GANs can produce realistic mouse input that stands up well to statistical and visual analysis but advanced discriminators can still detect synthetic trajectories.

CHAPTER 7

Limitations

The research shows that Generative Adversarial Networks (GANs) can produce synthetic mouse movement trajectories which match human input but several study limitations restricted its depth and scope.

7.1 Training Time and Resource Constraints

Training GANs is computationally intensive and time-consuming. The training process of each run needs extensive GPU time that spans multiple hours or days which restricts the number of experiments that can be conducted within a brief research period. The six-week experimental time frame proved too short to investigate multiple hyperparameter settings and alternative architectures and training configurations.

7.2 Limited Hyperparameter Exploration

Due to the resource and time constraints, the project was restricted to a set of generator and discriminator architectures, learning rates, latent dimensions and loss function settings. A more extensive hyperparameter tuning or the implementation of more advanced architectures such as attention-based models or TimeGAN could have achieved more realistic and temporally coherent results. However, this was not feasible within the time frame of the project.

7.3 Mode Collapse and Instability

GANs are known for their instability during training, the most common problem often being *mode collapse*. Mode collapse is where the generator starts producing repetitive or overly similar outputs, failing to represent the diversity of the training data. In practice, once a GAN reaches a mode-collapsed state, training often needs to be restarted from scratch, resulting in further time and resource costs. Several experiments in this study

encountered such collapse, limiting the ability to evaluate longer or more refined training runs.

7.4 Evaluation Scope

The evaluation included statistical similarity, classifier confusion assessment as well as visual inspection through t-SNE and UMAP. But real-time deployment, user study evaluations and comparison against adversarially robust detection models were beyond the project scope. The practical utility of the generated data in real-world systems remains partially untested because of this.

7.5 Dataset Limitations

The model was trained on a single dataset (Bogazici Mouse Dynamics), which, may not capture the full range of real-world mouse movement behaviors. The generator's ability to generalize to different user groups, hardware configurations, or tasks was not investigated.

7.6 Theoretical vs. Practical Performance Gap

Theoretically GANs can generate synthetic data that matches real data perfectly. The achievement of this level of performance in practice needs extensive training time and careful tuning and often requires advanced techniques to address training instability. The project demonstrates that GANs can generate realistic samples but practical challenges make it difficult to achieve theoretical perfection especially when generating complex temporal data.

CHAPTER 8

Future Work

8.1 Model Architecture Improvements

The current architecture relies on a WGAN-GP framework with LSTM-based Generator and Discriminator networks. The sequential data modeling capabilities of LSTMs do exist but they may struggle to detect both long-term dependencies and behavioral fine-grained variations. Future research could investigate more advanced architectures such as:

- **Transformer-based GANs** for capturing long-range temporal dependencies.
- **Attention mechanisms** to allow the generator to dynamically focus on past movement context.
- **Hybrid architectures** combining CNNs and RNNs for multi-resolution temporal learning.

8.2 Exploring TimeGAN for Temporal Realism

Future research could investigate the application of TimeGAN, a generative model that specializes in time-series data generation. TimeGAN differs from standard GANs because it combines adversarial training with supervised sequence modeling to maintain the temporal relationships found in the data. The generator benefits from this combined method to create sequences which maintain both statistical accuracy and temporal consistency.

TimeGAN may offer a promising solution for generating data that can deceive classifiers including LSTMs because mouse movement data consists of sequential patterns that depend heavily on timing and rhythm. The architecture of TimeGAN may solve the issues found in this research.

8.3 Training Optimization and Stability

GAN training is notoriously unstable, and results show signs of overfitting or mode collapse at later stages. Future research could explore the following to improve generalization and convergence:

- Hyperparameter tuning (e.g., learning rates, batch sizes, update frequency).
- Incorporation of training stabilization techniques (e.g., spectral normalization, progressive growing).
- Automated methods such as Bayesian optimization for architecture and training configuration search.

8.4 Adversarial Feedback with Stronger Discriminators

The current discriminator was designed to evaluate realism, but it may not be sufficient to challenge the generator in learning deeper behavioral nuances. Future work could:

- Use adversarial training for the generator from a pre-trained non-static deep LSTM classifier.
- Use ensemble discriminators, each focusing on different trajectory aspects (e.g., curvature, timing).

8.5 Expanded Evaluation Protocols

The evaluation process included statistical similarity, classifier confusion and visualization of latent structures. Additional evaluation would strengthen the validation of data realism, this could be done by:

- Human perception studies at a larger scale to determine how realistic the generated data seems to end-users.
- Real-world deployment or simulation environments to test whether generated inputs trigger detection in live systems.
- Task-specific benchmarks (e.g., performance of classifiers trained only on synthetic data).

8.6 Dataset Diversity and Generalization

The model was trained on a single dataset (Bogazici), this may restricts its ability to recognize different user types and input contexts. Future work could include:

- Training on multi-source datasets with diverse user behavior and screen resolutions.

8.7 Ethical Considerations and Responsible Use

As the realism of synthetic behavioral data improves, it creates new opportunities for its improper utilization. Future work should address:

- Guidelines for ethical use of synthetic human-like data in security research.
- The implementation of watermarking or detectability markers within generated data should occur to prevent abuse.

CHAPTER 9

Conclusion

This research investigated the generation of mouse movement trajectories through GANs to mimic authentic human behavior. The research developed from increasing security and usability testing requirements for behavioral biometrics and the technical and ethical problems associated with acquiring extensive human mouse movement datasets.

The WGAN-GP network with recurrent neural networks (LSTMs) in Generator and Discriminator components served to address this issue. The model received training from the Bogazici Mouse Dynamics dataset to generate sequences of (x, y, deltaT) coordinates. The objective included both the reproduction of statistical elements from actual trajectories and human movement characteristics including time-based dependencies and behavioral irregularities.

The GAN demonstrated effective capabilities to generate trajectories that both statistically matched authentic mouse behavior and appeared natural. The spatial and temporal Wasserstein distance metrics were low while qualitative visualizations of the generated paths showed natural appearances. Shallow SVM classifiers achieved near-random accuracy when evaluating mid-training outputs because the models failed to distinguish between generated and actual data at specific epochs. A deep LSTM-based classifier successfully distinguished real input data from generated data with high accuracy because it detected remaining imperfections in the temporal dynamics of the generated trajectories. Dimensionality reduction techniques (t-SNE and UMAP) showed that the generator excelled at preserving local structure but failed to achieve full distribution convergence with real data.

GANs demonstrate high-quality behavioral data generation capabilities but deep temporal dependencies between data points present a continuing difficulty. Although GANs possess theoretical capabilities to replicate data perfectly, there are often practical limitations that prevent them from achieving this ideal, especially when modeling complex temporal behavior like human mouse movement. The evaluation methods showed how deep models remain highly sensitive to minimal behavioral indicators while also proving the necessity of robust assessment approaches. This research establishes a base for developing future work on artificial behavioral data generation.

The investigation could continue by implementing TimeGAN alongside transformer-based models as well as stronger adversarial discriminators. The research should proceed by increasing dataset diversity together with improving training stability. Realistic synthetic mouse trajectory generation serves as a tool to enhance privacy protection through data augmentation and adversarial testing while enabling user behavior simulations in multiple practical fields.

REFERENCES

- [1] Alejandro Acien, Aythami Morales, Julian Fierrez, and Ruben Vera-Rodriguez. BeCAPTCHA-Mouse: Synthetic mouse trajectories and improved bot detection. *Pattern Recognition*, 127:108643, March 2022. Presents a mouse-based bot detection system using neuromotor modeling and synthetic trajectory generation. Public code available at: <https://github.com/BiDALab/BeCAPTCHA-Mouse>.
- [2] Ahmed Awad E Ahmed and Issa Traore. A new biometric technology based on mouse dynamics. *IEEE Transactions on dependable and secure computing*, 4(3):165–179, 2007.
- [3] Santiago Escuder Folch. Web Bot Detection Using Mouse Movement. Master’s thesis, Universitat Politècnica de Catalunya, October 2022. Master’s thesis exploring the use of LightGBM and LSTM models on real and synthetic mouse movement datasets.
- [4] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.
- [5] Christos Iliou, Theodoros Kostoulas, Theodora Tsikrika, Vasilis Katos, Stefanos Vrochidis, and Ioannis Kompatsiaris. Detection of advanced web bots by combining web logs with mouse behavioural biometrics. *Digital threats: research and practice*, 2(3):1–26, 2021.
- [6] Christos Iliou, Theodoros Kostoulas, Theodora Tsikrika, Vasilis Katos, Stefanos Vrochidis, and Ioannis Kompatsiaris. Web bot detection evasion using generative adversarial networks. In *2021 IEEE International Conference on Cyber Security and Resilience (CSR)*, pages 115–120. IEEE, 2021.
- [7] Abdul Jabbar, Xi Li, and Bourahla Omar. A survey on generative adversarial networks: Variants, applications, and training. *ACM Computing Surveys (CSUR)*, 54(8):1–49, 2021.
- [8] Simon Khan, Charles Devlen, Michael Manno, and Daqing Hou. Mouse dynamics behavioral biometrics: A survey. *ACM Computing Surveys*, 56(6):1–33, 2024.

- [9] Arjen Aykan Kılıç, Metehan Yıldırım, and Emin Anarım. Bogazici mouse dynamics dataset. *Data in Brief*, 36:107094, 2021.
- [10] Jinwook Kim, Kyungroul Lee, and Hanjo Jeong. Real-time mouse data protection method using gans for image-based user authentication based on getcursorpos () and setcursorpos () functions. *Applied Sciences*, 15(2):977, 2025.
- [11] Hongfeng Niu, Ang Wei, Yunpeng Song, and Zhongmin Cai. Exploring visual representations of computer mouse movements for bot detection using deep learning approaches. <https://ssrn.com/abstract=4329684>, January 2023. Preprint. Proposes a CNN-based method using augmented trajectory images to detect bots via mouse movements.
- [12] Vasvi Sharma and Siddhartha Chauhan. Web bot detection using keyboard behavioural analysis. *AIJR Proceedings*, pages 79–90, 2025.
- [13] Ang Wei, Yuxuan Zhao, and Zhongmin Cai. A deep learning approach to web bot detection using mouse behavioral biometrics. In *Biometric Recognition: 14th Chinese Conference, CCBR 2019, Zhuzhou, China, October 12–13, 2019, Proceedings 14*, pages 388–395. Springer, 2019.
- [14] Roman V Yampolskiy and Venu Govindaraju. Behavioural biometrics: a survey and classification. *International Journal of Biometrics*, 1(1):81–113, 2008.