Subject identification through Neural SDEs on human gaze trajectories

Natural Interaction and Affective Computing

Mattia Vinciguerra and Ivan Coppola (45962A, 46647A)

September 29, 2025

UNIVERSITÀ
DEGLI STUDI
DI MILANO

Table of Contents

- Biometric identification through gaze trajectories
- How eye-movements reflecting behavioral and cognitive patterns
- **Challenges**: stochastic, variable-length trajectories influenced by perception, cognition and micro eye-movements
  - Randomness in gaze: neural noise, perceptual variability, environmental factors
- Stochastic Differential Equations (SDEs): drift + diffusion to describe temporal dynamics
- **Neural SDEs**: parameterize drift and diffusion via neural networks
- Results: stochastic deep models capture dynamics, accuracy above baseline
- Applications: biometrics, human–computer interaction, neurological assessment

Biometrics

- Definition: authentication based on physical and/or behavioral characteristics
- Rapidly expanding research field for security and authentication
  - Widely adopted in daily life (fingerprint, face recognition on smartphones, laptops, personal devices)
- Traditional biometrics: fingerprints, face, iris, retina
- Eye movements as biometric traits:
  - Strongly linked to cognitive processes
  - Hard to imitate
  - Based on unique internal behavioral patterns
- Less intrusive compared to physical biometrics

Human gaze trajectories

- Gaze trajectories: continuous paths of the eye while exploring the environment
- Derived from eye-tracking devices capturing non-verbal and unconscious responses
- Reflect interplay between sensory (bottom-up) and cognitive (top-down) mechanisms
- Composed of:
  - **Fixations**: periods of relative stability
  - **Saccades**: rapid movements between fixations
  - Additional events: blinks, smooth pursuit

- Defined in two main ways:
  - **Functional**: maintaining target on the fovea, enabling visual processing
  - **Oculomotor**: periods when the eye remains relatively still
- Functional view: achieved via microsaccades, drift, tremor
- Oculomotor view: emphasizes kinematic stability of eye position
- Ambiguity arises from different frames of reference (eye vs. head vs. environment)

Saccades

1 Introduction

- Rapid eye movements that redirect gaze to a new region of the scene
- Typically occur between fixations
- Definitions vary depending on fixation interpretation:
  - **Functional**: saccades are distinct events, fixations include all stabilization (even pursuit, VOR)
  - **Oculomotor**: saccades are intervals between periods of relative eye stillness
- Role: enable exploration of the environment by shifting the fovea onto new regions of interest

Individual uniqueness

- **Ideal biometric traits:** unique, stable, universal, easy to measure, cost-effective, acceptable
- No single biometric satisfies all criteria:
  - Fingerprints/retina: highly distinctive but intrusive
  - Voice/face: easy to capture but less discriminative
- **Gaze as biometric:**
  - Gaze trajectories are partly random, partly guided by cognition
  - Reflect personal adaptation and unique internal models
  - Each individual shows distinctive fixation patterns
- **FIFA Dataset:**
  - Eye-tracking on natural images, same for all participants
  - Originally for predicting gaze locations
  - Our use: identify individuals by unique individual fixation patterns

- **Ecological individuality:** unique combination of traits, interactions, and behaviours
- **Foraging behaviour:** animals balance search, consumption, and relocation
  - Exploration–exploitation cycle
  - Stable inter-individual differences (personality, specialization)
- **From food to information:** ancestral foraging strategies evolved into cognitive information foraging
- **Gaze as foraging:**
  - Saccades = exploration (relocation)
  - Fixations = exploitation (local resource use)
  - Eye movements shaped by stimuli, tasks, and internal states
- **Implication:** gaze trajectories reveal stable, individual-specific patterns analogous to foraging strategies

Ornstein–Uhlenbeck process

- **OU process:** mean-reverting Gaussian Markov process
- Linear SDE in $\mathbb{R}^d$:

$$d\mathbf{z}(t) = -\boldsymbol{\Theta}(\mathbf{z}(t) - \boldsymbol{\mu})\, dt + \boldsymbol{\Sigma}^{1/2}\, d\mathbf{w}(t)$$

- Parameters:
  - $\boldsymbol{\Theta}$: mean-reversion strength
  - $\boldsymbol{\mu}$: equilibrium (fixation point)
  - $\boldsymbol{\Sigma}$: diffusion covariance (micro-saccades, tremor, drift, noise)
  - $\mathbf{w}(t)$: standard Brownian motion
- Properties:
  - Stationary, Gaussian, ergodic
  - Closed-form transition densities
- **Inductive biases for gaze modeling:**
  - Mean reversion around fixation point
  - Temporally correlated Gaussian perturbations
  - Continuous-time dynamics for irregular sampling

Neural Stochastic Differential Equations (Neural SDEs)

- **Neural SDEs:** extension of SDEs where drift and diffusion are parameterized by neural networks.

$$d\mathbf{z}(t) = f_\theta(\mathbf{z}(t), t)\, dt + g_\phi(\mathbf{z}(t), t)\, d\mathbf{w}(t),$$

  where:
  - $\mathbf{z}(t) \in \mathbb{R}^d$ is the latent state at time t,
  - $f_\theta(\mathbf{z}, t)$ is the **drift** function, parameterized by neural network with parameters θ,
  - $g_\phi(\mathbf{z}, t)$ is the **diffusion** function, parameterized by neural network with parameters ϕ,
  - $\mathbf{w}(t)$ is standard Brownian motion (Wiener process).
- Applications: finance, particle systems, population dynamics, genetics
- **Motivation for gaze modeling:**
  - Fixations: small, stable micro-movements around targets
  - Trajectories are stochastic, variable in length, influenced by perception and cognition
  - **Continuous-time dynamics**: discrete-time models (e.g., RNNs) struggle with irregular sampling

Table of Contents

Methodology Overview

- **Objective**: Subject identification from gaze dynamics using Neural SDE-based architectures
- **Approach**: Task-independent eye movement biometrics
- **Dataset**: Free-viewing task with raw gaze data only
- **Models**: Baseline sequence models vs. Neural differential equations
- **Focus**: Assess whether stochastic diffusion captures user-specific patterns

Dataset: Fixations In Faces (FiFa)

- **Source**: Moran Cerf et al. - FiFa database
- **Task**: Free-viewing of images (2 seconds per image)
- **Participants**: 8 subjects
- **Data**: 200 images per subject = 200 scanpaths per subject
- **Structure**: Each scanpath = sequence of 2021 gaze (x, y) coordinates

Preprocessing Pipeline

- **Event classification**: I-VT algorithm (Velocity-Threshold Identification)
- **Segmentation**: Raw gaze $\rightarrow$ fixations + saccades
- **Filtering**: Retain only fixation events for modeling
- **Centering**: Zero-center fixations by subtracting scanpath mean
- **OU alignment**: Centered process aligns with Ornstein-Uhlenbeck prior

$$d\mathbf{z}_t = -\,\mathbf{z}(t)\,dt + {}^{1/2}\,d\mathbf{w}(t)$$

Baseline Models

- **RNN**: Basic recurrent unit with $\mathbf{h}_t = \phi(\mathbf{W}_{hh}\mathbf{h}_{t-1} + \mathbf{W}_{xh}\tilde{\mathbf{x}}_t + \mathbf{b}_h)$
- **GRU**: Gated recurrent unit with update/reset gates

$$\mathbf{h}_t = (1 - \mathbf{z}_t) \odot \mathbf{h}_{t-1} + \mathbf{z}_t \odot \tilde{\mathbf{h}}_t$$

- **LSTM**: Long short-term memory with explicit cell state

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t$$

Neural Ordinary Differential Equations

- **Classical ODE**: Deterministic, Markovian, continuous-time, smooth
- **Neural extension**: Replace parametric drift with neural network
- **Formulation**:

$$\frac{d\mathbf{z}(t)}{dt} = f(\mathbf{z}(t), t; \theta), \quad \mathbf{z}(0) = g(\mathbf{x}_{1:T})$$

- **Advantages**: Memory-efficient adjoint training, adaptive computation
- **Challenges**: Purely deterministic, may underfit noisy gaze data

Neural Controlled Differential Equations

- **Extension**: ODEs driven by external control path
- **Control-driven dynamics**: Evolution depends on input trajectory velocity
- **Formulation**:
$$\frac{d\mathbf{z}(t)}{dt} = f(\mathbf{z}(t), t)\frac{d\mathbf{x}(t)}{dt}$$
- **Advantages**: Handles irregular sampling via spline interpolation
- **Applications**: Sequential data with nonuniform time points

Neural Stochastic Differential Equations

- **Extension**: Add stochastic diffusion term to capture uncertainty
- **Neural parameterization**: Both drift and diffusion are neural networks
- **Formulation**:
$$d\mathbf{z}(t) = f(\mathbf{z}(t), t; \theta_f)\, dt + g(\mathbf{z}(t), t; \theta_g)\, d\mathbf{w}(t)$$
- **Components**: Drift f (deterministic) + Diffusion g (stochastic)
- **Goal**: Model variability in human eye movement patterns

Stable Neural SDE Variants

- **Problem**: Generic Neural SDEs can be unstable
- **Solution**: Three stable architectures (Oh et al. 2024)
- **Langevin-type (LSDE)**: $d\mathbf{z}(t) = \gamma(\mathbf{z}(t); \theta_\gamma)\, dt + \sigma(t; \theta_\sigma)\, d\mathbf{w}(t)$
  - Drift function not explicitly dependent on time
- **Linear Noise (LNSDE)**: $d\mathbf{z}(t) = \gamma(t, \mathbf{z}(t); \theta_\gamma)\, dt + \sigma(t; \theta_\sigma)\mathbf{z}(t)\, d\mathbf{w}(t)$
  - Diffusion term scales linearly with the latent state
- **Geometric (GSDE)**: $\frac{d\mathbf{z}(t)}{\mathbf{z}(t)} = \gamma(t, \mathbf{z}(t); \theta_\gamma)\, dt + \sigma(t; \theta_\sigma) d\mathbf{w}(t)$
  - Positive constraint and 0 state absorbing

Controlled Path Enhancement

- **Motivation**: Better capture sequential dependencies
- **Augmented state**: $\overline{\mathbf{z}}(t) = \zeta(t, \mathbf{z}(t), X(t); \theta_\zeta)$
- **Control path**: $X(t)$ = cubic spline interpolation of scanpath
- **Integration**: Replace $\mathbf{z}(t)$ with $\overline{\mathbf{z}}(t)$ in drift functions
- **Benefit**: Drift depends on current state + temporal history
- **Limitation**: Model becomes purely discriminative (no generation)

Training and Evaluation

- **Classification head**: Linear layer + softmax on final state $\mathbf{h}_T$
- **Loss function**: Cross-entropy $\mathcal{L} = -\sum_{i=1}^{N} y_i \log \hat{y}_i$
- **Hyperparameter tuning**: Grid search over batch size + hidden dimension
- **Metrics**:
  - Epochs to convergence (validation loss minimum)
  - Cross-entropy loss
  - Fixation-level accuracy (individual fixations)
  - Scanpath-level accuracy (majority voting per sequence)

Table of Contents

3 Results and Conclusions

Performance Comparison

Table: Model performance metrics (**best**, *second best*)

Model	Epoch	Val Loss	Fix. Acc.	Scan. Acc.
RNN	19	1.970	0.255	0.312
GRU	33	1.500	0.502	0.704
LSTM	35	1.580	0.489	0.675
CDE	92	1.900	0.323	0.433
ODE	33	1.990	0.212	0.267
SDE	45	1.790	0.355	0.451
LSDE	165	*1.270*	**0.583**	**0.787**
LNSDE	169	**1.217**	*0.571*	*0.775*
GSDE	17	1.960	0.265	0.342

Key Findings: Baseline Models

- **GRU dominance**: Best among classical sequence models ($\sim$50% fix. acc., $\sim$70% scan. acc.)
- **LSTM performance**: Competitive but slightly behind GRU
- **RNN limitation**: Poorest performance among baselines
- **Conclusion**: Gating mechanisms crucial for noisy gaze sequences

Key Findings: Continuous-Time Models

- **Neural ODE**: Poorest overall performance
  - Deterministic dynamics insufficient for eye movement variability
- **Neural CDE**: Moderate performance
  - Handles irregular sampling but underfits discriminative task
- **Generic Neural SDE**: Marginal improvement over ODE and CDE but under the baseline ($\sim$36% fix. acc., $\sim$45% scan. acc.)
  - Stochastic modeling is necessary for capturing eye movement variability but noise alone is insufficient

Key Findings: Stable Neural SDEs

- **LSDE (Langevin-type)**: Best overall performance ($\sim$58% fix. acc., $\sim$79% scan. acc.)
  - Mean-reverting dynamics align with fixation behavior
- **LNSDE (Linear Noise)**: Second best performance
  - State-dependent diffusion captures variability
- **GSDE (Geometric)**: Underperforms stable variants
  - Constraints misaligned with our data

Limitations and Future Work

- **Current limitations**:
  - Controlled path models are purely discriminative
  - Cannot generate novel gaze sequences
  - Limited to fixation-based analysis
- **Future directions**:
  - Generative extensions for sequence synthesis
  - Saccade-based classification benchmarks

Final Conclusions

- **Main finding**: Stable Neural SDEs provide powerful framework for gaze-based identification
- **Practical impact**: Task-independent biometrics from raw gaze data
- **Broader implications**: Applicable to other irregular time series domains

Subject identification through Neural SDEs on human gaze trajectories

Thank you for listening