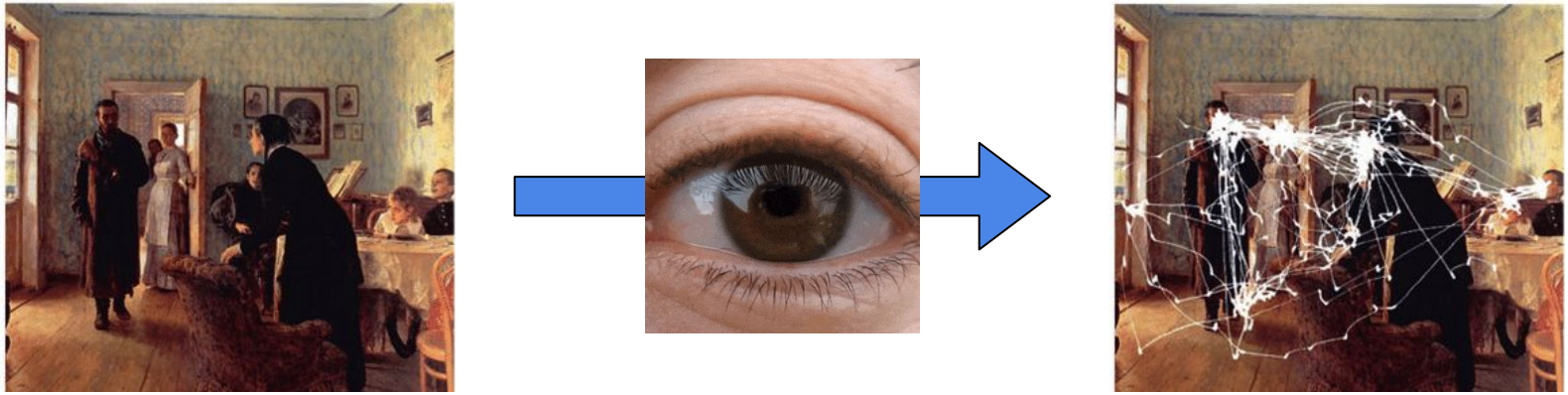
Modeling Human Eye Movements with Neural Networks in a Maze-Solving Task

Jason Li, Nicholas Watters, Yingting (Sandy) Wang, Hansem Sohn, Mehrdad Jazayeri

Dept. of Brain and Cognitive Sciences, Massachusetts Institute of Technology

NeurIPS 2022 Gaze Meets ML Workshop

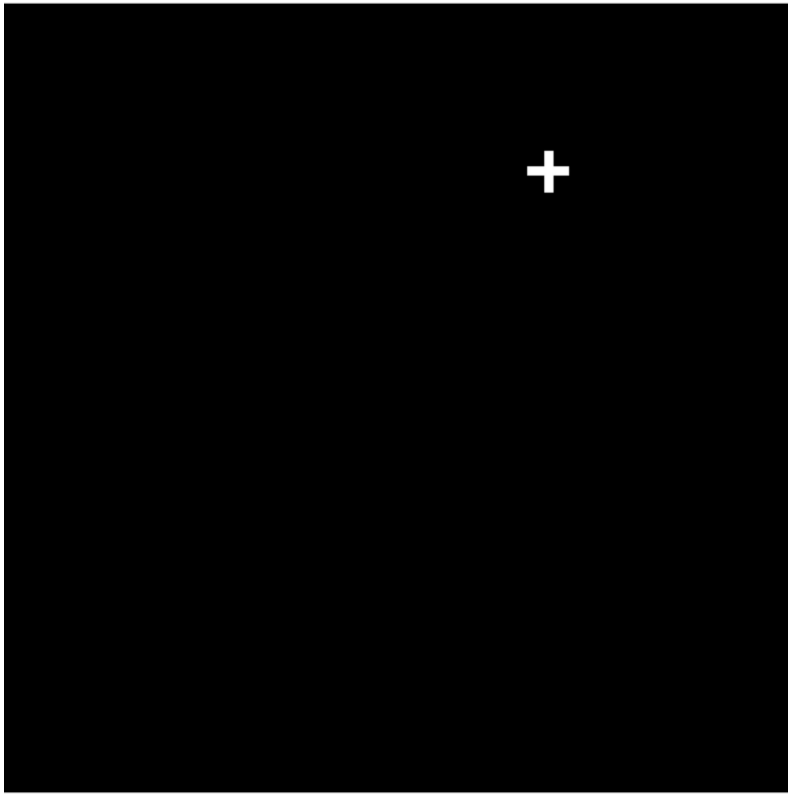
Dec. 3, 2022



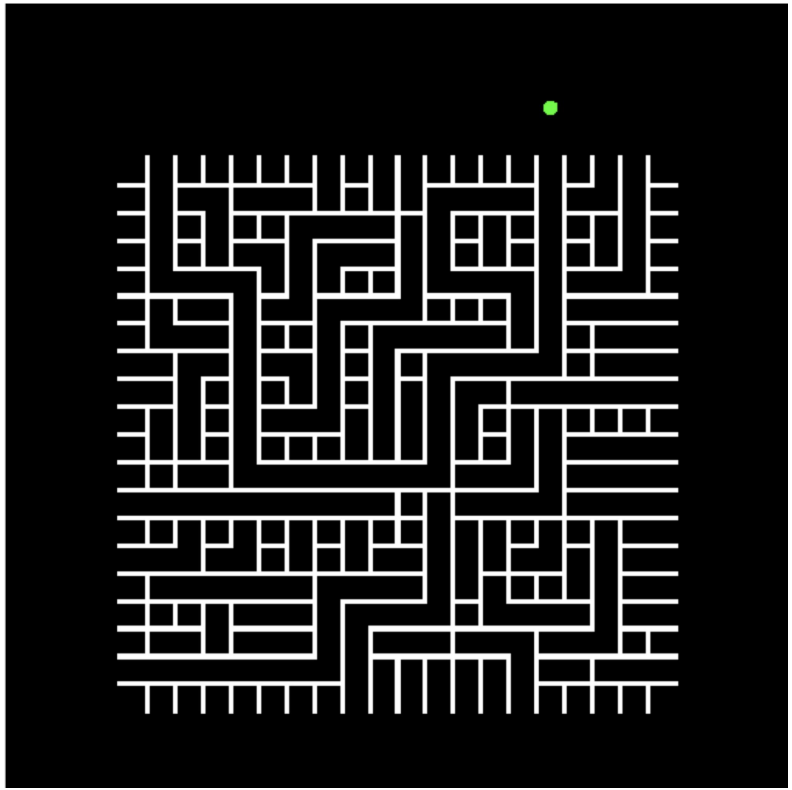
(Yarbus, 1967)

Goal: Build a **generative model of eye movements** that incorporates features of human vision like a fovea and discrete saccades.

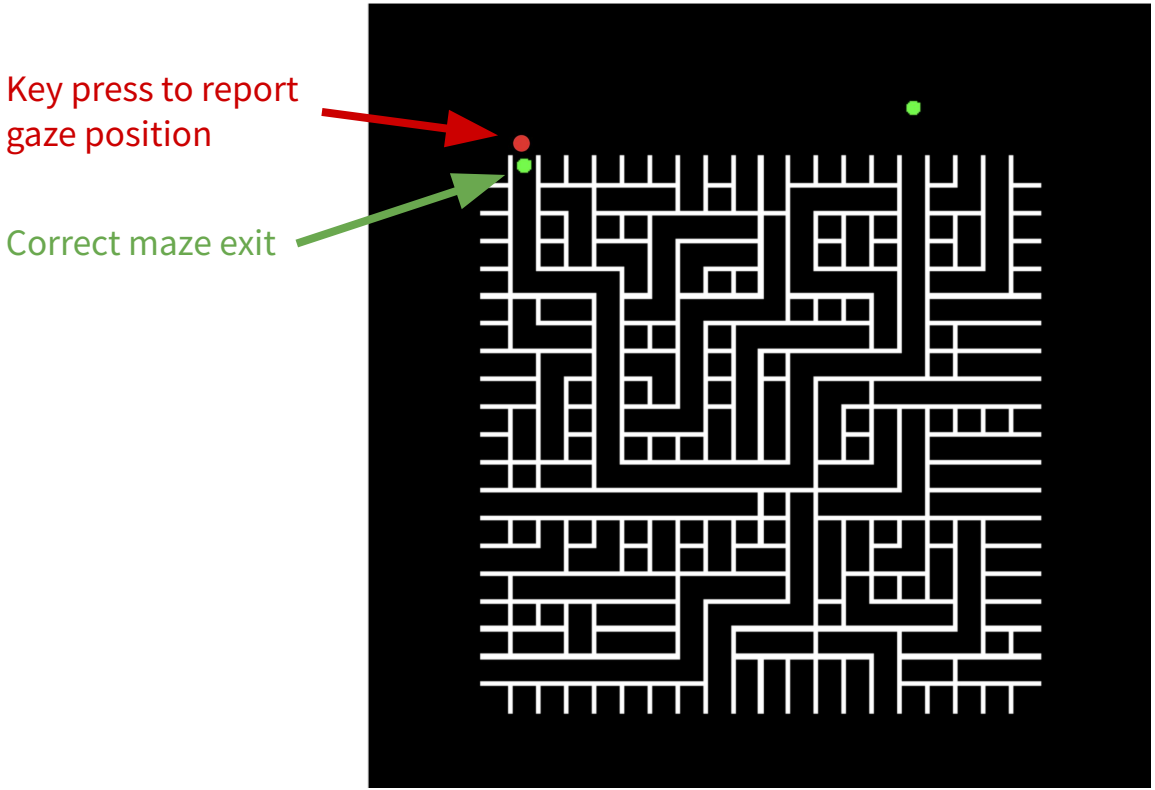
Maze task: Fixation at Entrance



Maze task: Maze Presentation

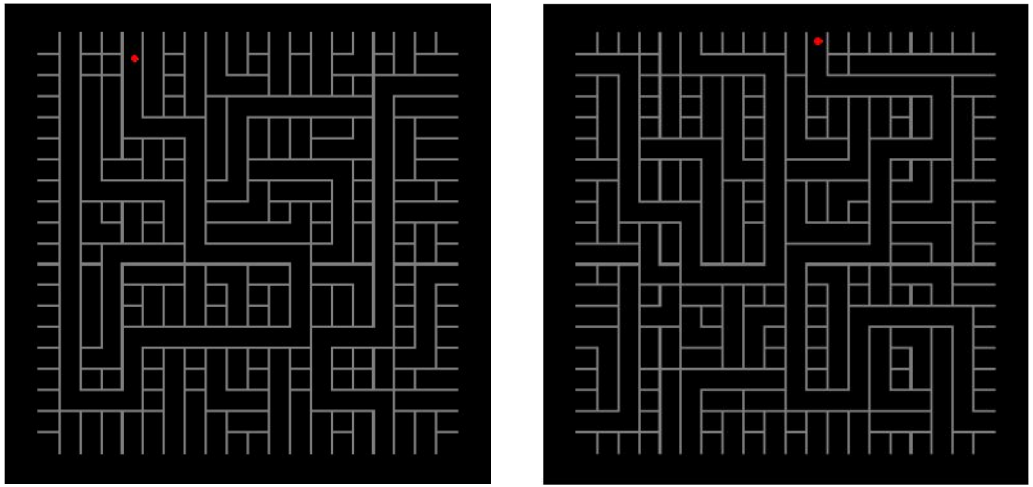


Maze task: Saccade to the Exit



Maze task

- Rich variety of spatial plans allows study of task-driven saccade sequences
- Eye movements are largely consistent across humans
- Allows us to test mental simulation as a strategy guiding eye movements



(videos shown 30% full speed)

Maze task

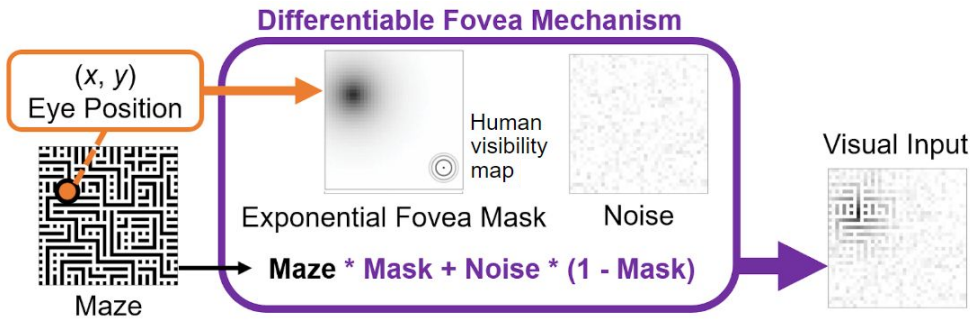
- Rich variety of spatial plans allows study of task-driven saccade sequences
- Eye movements are largely consistent across humans
- Allows us to test mental simulation as a strategy guiding eye movements

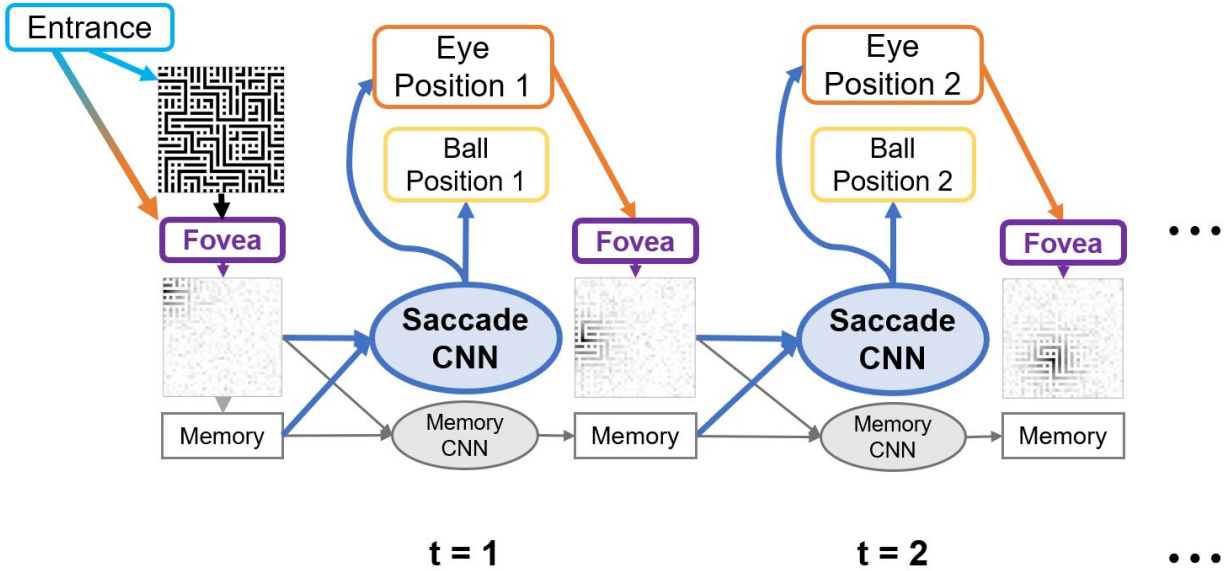
Human data collection

- 14 human volunteers
- Each subject completed ~400 trials
- Saccades extracted by filtering and thresholding eye velocity

Gaze RNN Models

- Novel differentiable fovea mechanism allows end-to-end training
- ConvRNN architecture unrolls through time for any number of saccades





3 Gaze RNN Variations

- **EXIT**: Reach exit in as few saccades as possible

Minimize $L_{\text{EXIT}} = \frac{1}{n}\sum_{i=1}^{n}(\hat{p}_i^{\text{eye}} - p^{\text{exit}})^2$

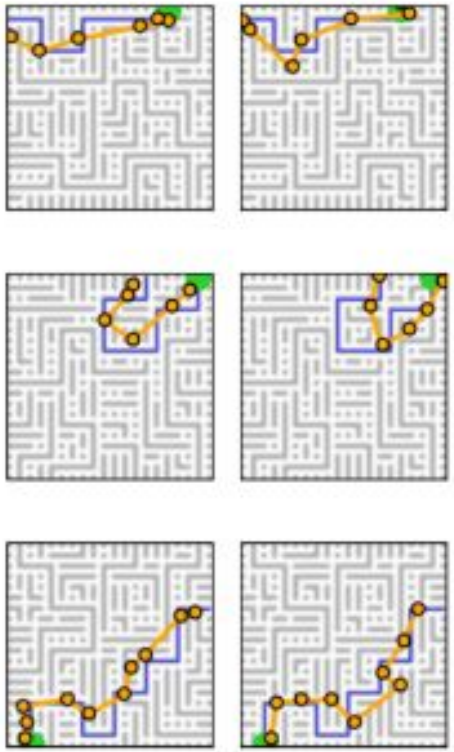
- **SIM**: Track an imaginary “ball” moving through the maze

Minimize $L_{\text{SIM}} = \frac{1}{n}\sum_{i=1}^{n}(\hat{p}_i^{\text{ball}} - p_i^{\text{ball}})^2$

- **HYBRID**: Weighted sum of the two loss terms

Minimize $L_{\text{HYBRID}} = \beta \cdot L_{\text{EXIT}} + (1-\beta) \cdot L_{\text{SIM}}$

Results: Qualitative Behavior



Human 1 Human 2

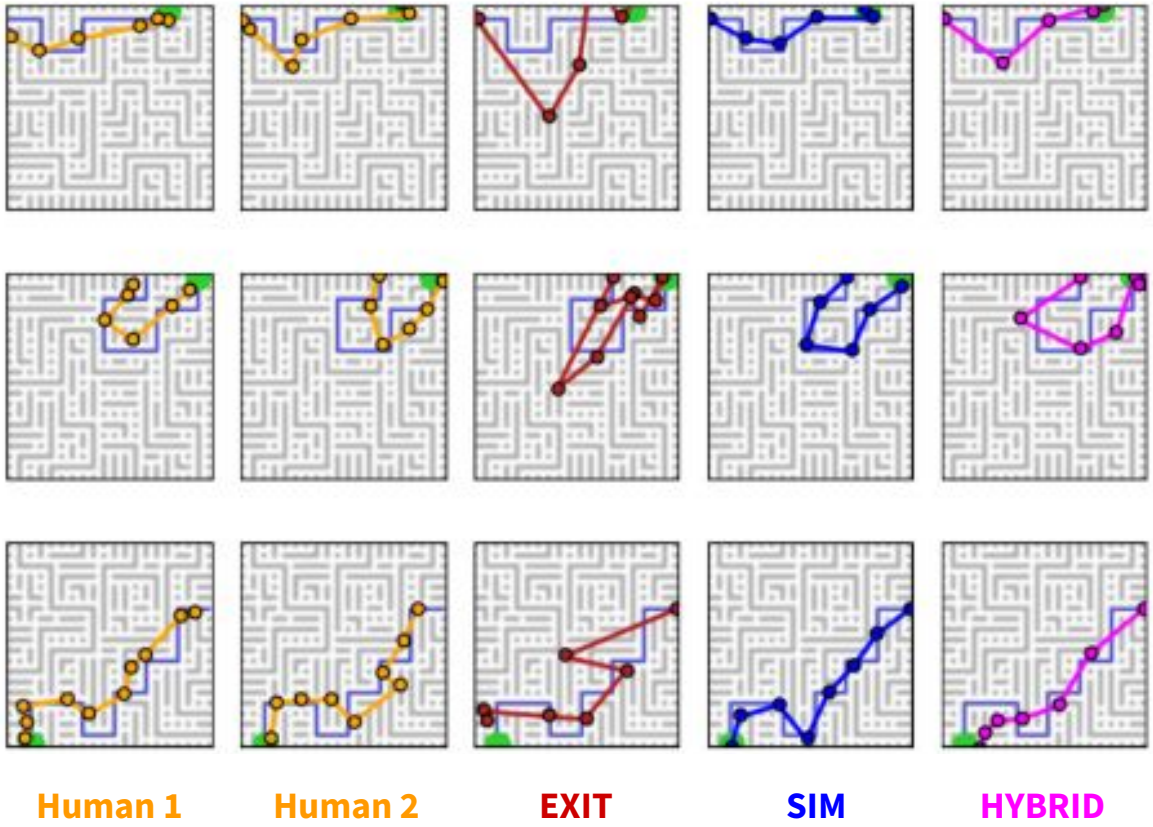
Exit

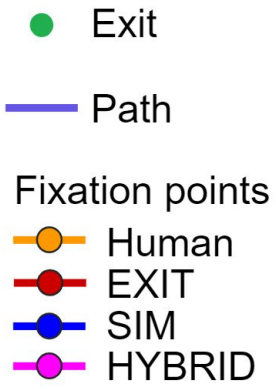
Path

Fixation points

Human

Results: Qualitative Behavior

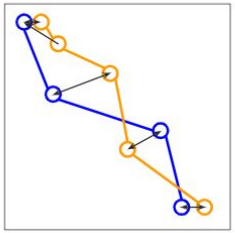




Results: Quantitative Metrics

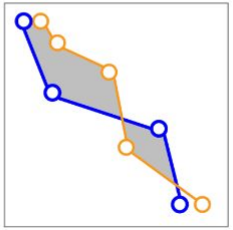
Nearest Neighbors Distance

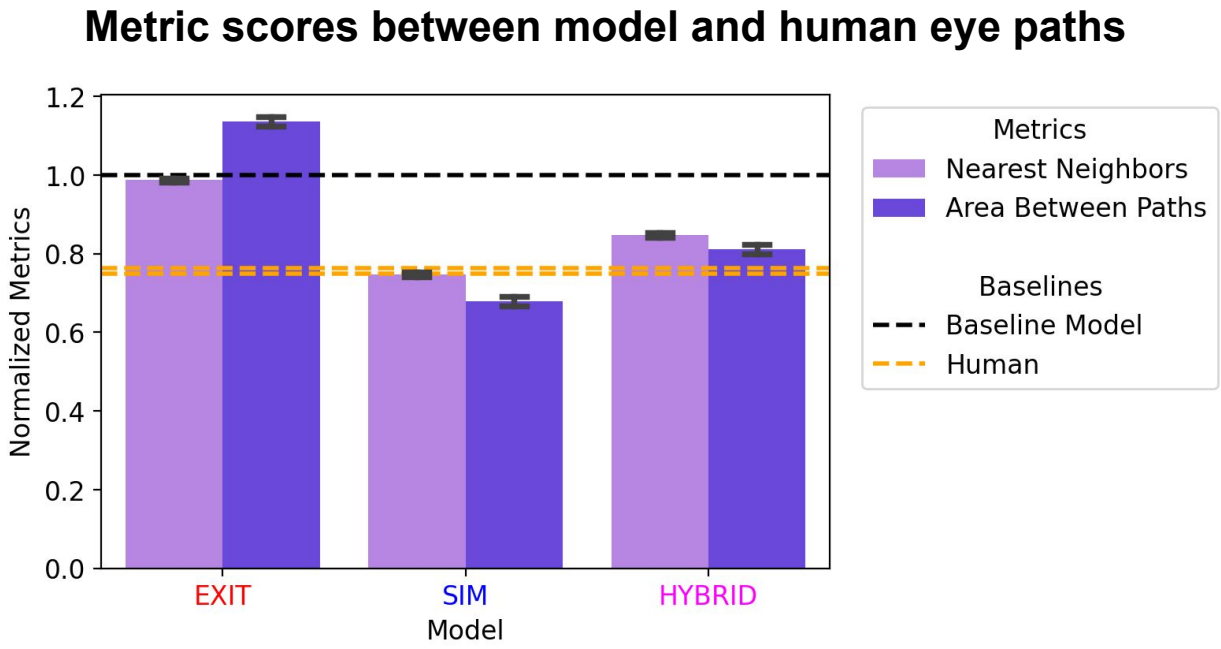
Mean of the nearest point in path **A**
to every point in path **B** and vice versa



Area Between Paths

Total plane area of the polygon(s)
formed between paths **A** and **B**





Conclusions & Future Work

- In the maze-solving task, a gaze RNN trained to run an internal simulation of a ball moving through the maze generates more human-like eye movements than a model trained to solve the task optimally
- Humans may employ a similar mental simulation when solving this task
- Future work may address:
  - Biological accuracy of model fovea hyperparameters
  - Introducing non-constant simulation speed
  - Allow gaze RNN to control fixation durations
  - Apply gaze RNNs to other task domains

Acknowledgements

Thank you!