



# Towards the Coordination of Eye, Body and Context in Daily Activities

Research Talk

---

Zhiming Hu

January 2, 2024

Perceptual User Interfaces Group, University of Stuttgart

Computational Biophysics and Biorobotics Group, University of Stuttgart

Stuttgart Center for Simulation Science (SimTech)

Personal Introduction

Research Statement



### Education Background

- Ph.D. in Computer Software and Theory 2017.09-2022.07  
**Peking University**, Supervised by Prof. Guoping Wang
- B.Eng. in Optical Engineering 2013.09-2017.07  
**Beijing Institute of Technology**

### Academic Positions

- Post-doctoral Researcher 2022.08-now  
**University of Stuttgart**, Led by Prof. Andreas Bulling & Prof. Syn Schmitt



- **Best Doctoral Student Paper Award Nominees at INTERACT 2023** (top 5%), 2023
- SimTech Postdoctoral Fellowship, 2022
- **National Scholarship** (top 2%), 2021
- **TVCG Best Journal Award Nominees at IEEE VR 2021** (top 2%, first time for Chinese researchers), 2021
- CSC (China Scholarship Council) Scholarship, 2020
- Chancellor's Scholarship (top 2%), 2020
- Leo KoGuan Scholarship (top 5%), 2019
- Leader Scholarship (top 0.2%, 7 out of over 3800 students), 2017
- **National Scholarship** (top 2%), 2016
- **National Scholarship** (top 2%), 2014



Personal Introduction

Research Statement



- Human-computer interaction
- Virtual reality
- Eye tracking
- Human-centred artificial intelligence

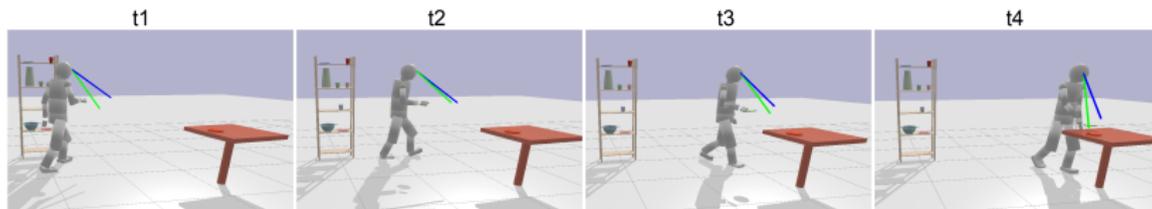
### Research goal

Develop deep learning methods for modelling human behaviours in activities of daily living



## Towards the Coordination of Eye, Body and Context in Daily Activities

- Eye and body movements are correlated in daily activities
- Eye and body movements are influenced by the context, e.g. *scene environment, action, and task*



Coordination of eye, body and context

### Towards the Coordination of Eye, Body and Context in Daily Activities

- Everyday Human Behaviour Sensing
- Computational Human Activity Analysis



### SGaze: An Eye-Head Coordination Model for Gaze Prediction

$$\tilde{x}_g = \alpha_x \cdot \tilde{v}_{hx}(t + \Delta t_x) + \beta_x \cdot a_{hx} + b_x \cdot x_S + c_x$$

$$\tilde{y}_g = \alpha_y \cdot \tilde{v}_{hy}(t + \Delta t_y) + b_y \cdot y_S + c_y$$

$\tilde{x}_g, \tilde{y}_g$ : predicted eye gaze

$\tilde{v}_{hx}, \tilde{v}_{hy}$ : head velocity

$\Delta t_x, \Delta t_y$ : time interval between gaze and head

$a_{hx}$ : horizontal head acceleration

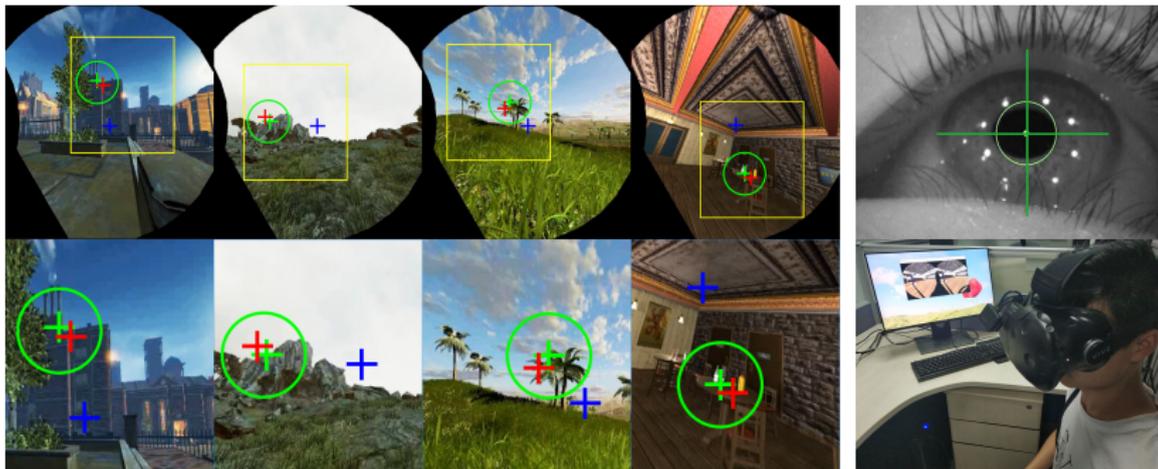
$x_S, y_S$ : salient positions

$\alpha_x, \alpha_y, \beta_x, b_x, b_y, c_x, c_y$ : learned parameters

[Hu TVCG'19]



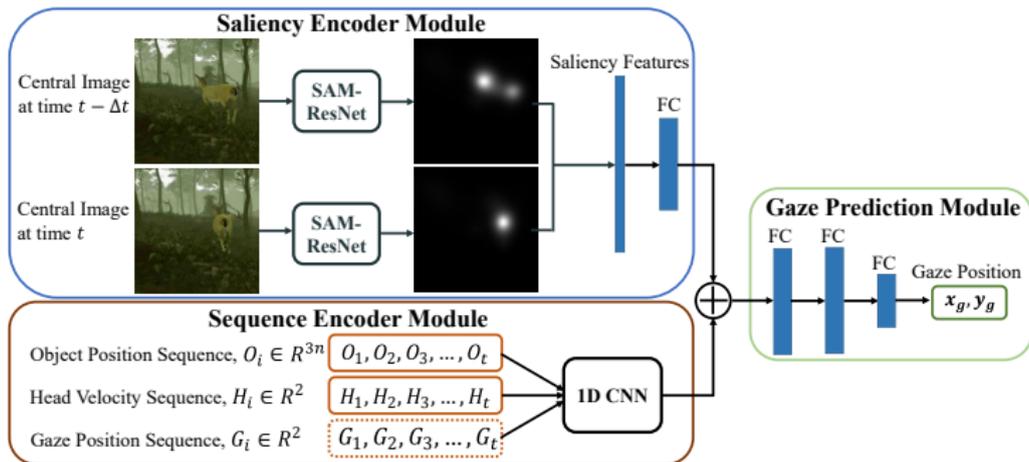
## SGaze: An Eye-Head Coordination Model for Gaze Prediction



[Hu TVCG'19]

## DGaze: CNN-based Gaze Prediction in Dynamic Scenes

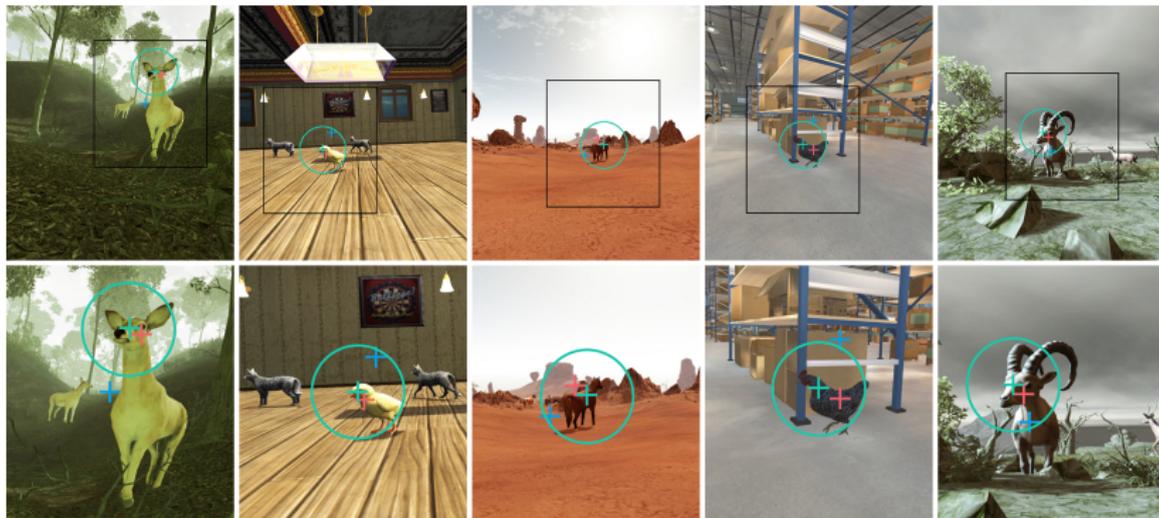
- Gaze estimation using VR content, and head movements
- Gaze forecasting using past gaze positions



[Hu TVCG'20]



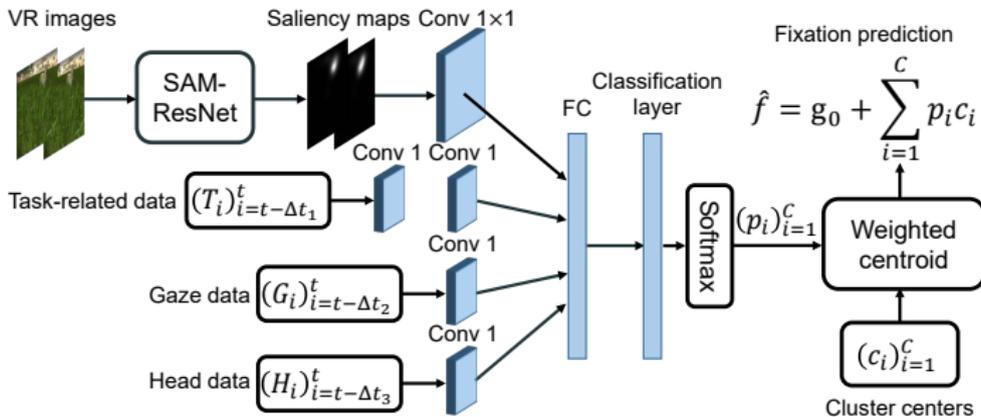
## DGaze: CNN-based Gaze Prediction in Dynamic Scenes



[Hu TVCG'20]

## FixationNet: Gaze Forecasting in Task-Oriented Environments

- Extract features from VR content, past gaze and head data
- Forecast fixation using prior knowledge of gaze distribution



[Hu TVCG'21 Best Journal Nominees Award]



## FixationNet: Gaze Forecasting in Task-Oriented Environments

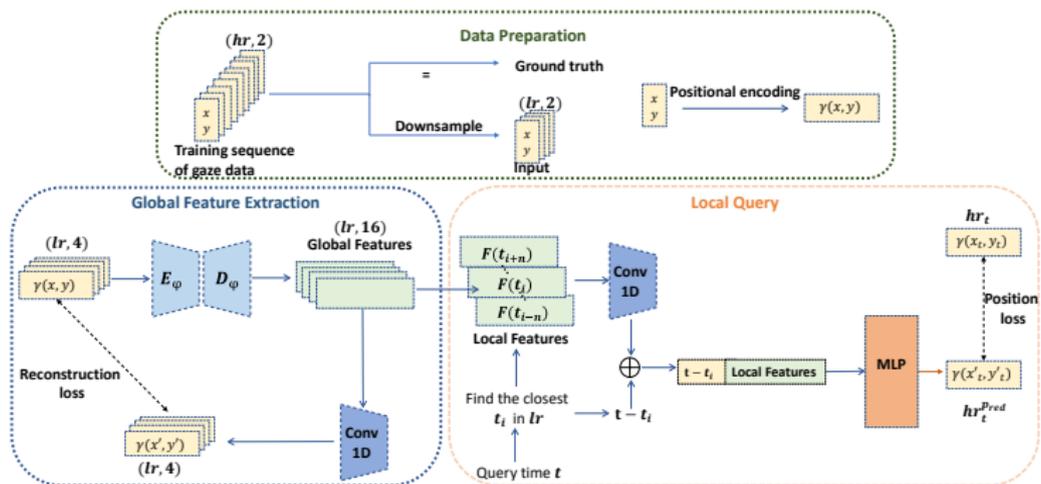
- Stimuli: immersive virtual environments
- Task: visual search



[Hu TVCG'21 Best Journal Award Nominees]

## SUPREYES: SUPER Resolution for EYES

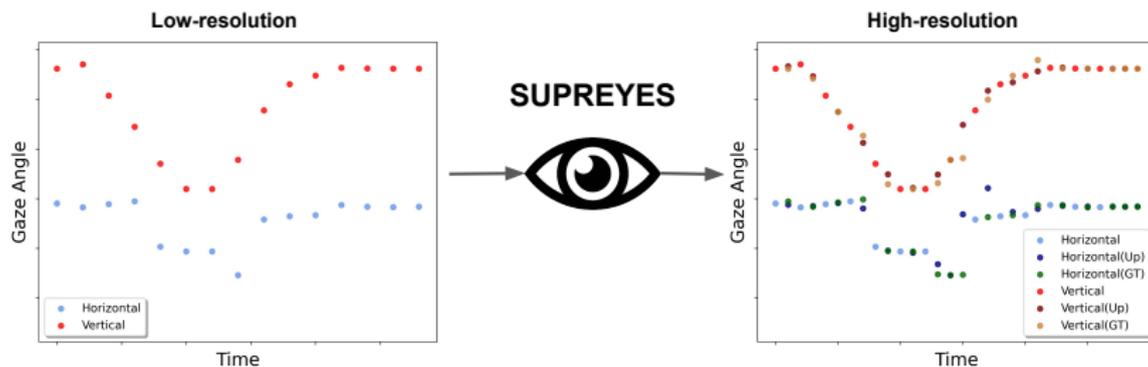
- Implicit neural representation learning
- Global feature extraction and local query



[jiao UIST'23]



## SUPREYES: SUPER Resolution for EYES

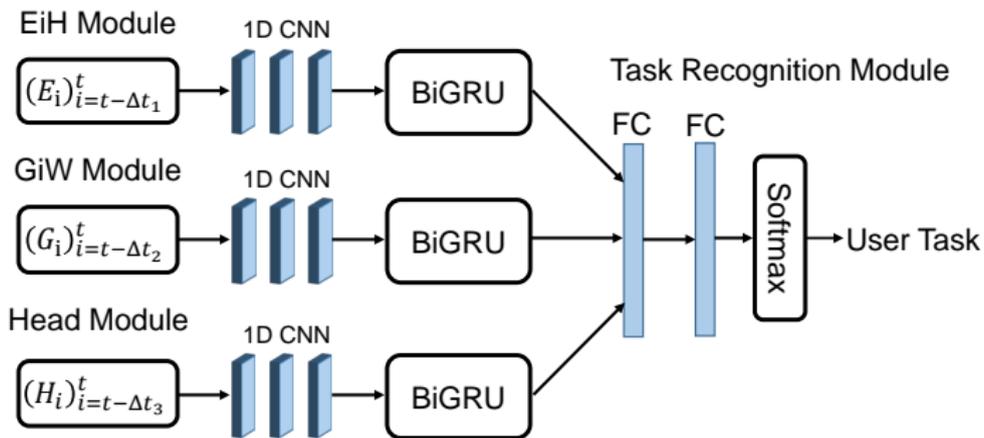


[Jiao UIST'23]



## EHTask: Task Recognition from Eye and Head Movements

- Extract features from eye and head movements
- Recognise user tasks from eye-head features



[Hu TVCG'22]



## EHTask: Task Recognition from Eye and Head Movements

Task recognition performances in cross-user and cross-scene settings

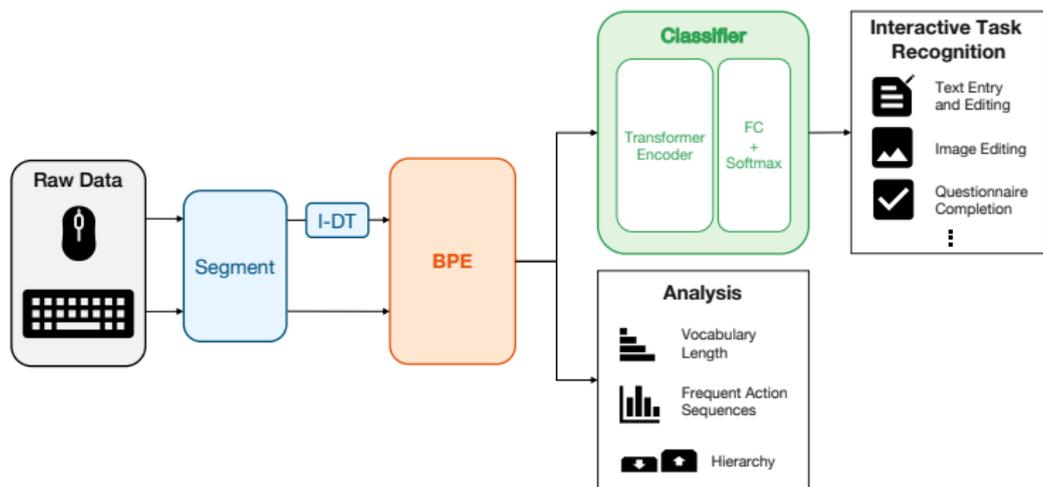
		Ours	<i>LDA_r</i>	<i>LDA_h</i>	<i>SVM_r</i>	<i>SVM_h</i>	<i>BC_r</i>	<i>BC_h</i>	<i>RFo_r</i>	<i>RFo_h</i>	<i>RFe</i>
Cross-User	Window	<b>84.4%</b>	37.2%	54.0%	29.5%	54.3%	41.5%	49.3%	<u>62.8%</u>	58.0%	48.7%
	MV	<b>97.8%</b>	42.8%	76.1%	34.2%	75.3%	47.5%	65.3%	83.1%	<u>88.9%</u>	68.3%
Cross-Scene	Window	<b>82.1%</b>	37.2%	53.8%	26.3%	54.1%	41.2%	49.0%	<u>62.6%</u>	57.9%	48.3%
	MV	<b>96.4%</b>	41.9%	74.2%	26.7%	75.3%	47.5%	64.4%	83.6%	<u>87.2%</u>	72.2%

Our method significantly outperforms prior methods in both cross-user and cross-scene settings



## Modelling Interactive Behaviour using NLP Methods

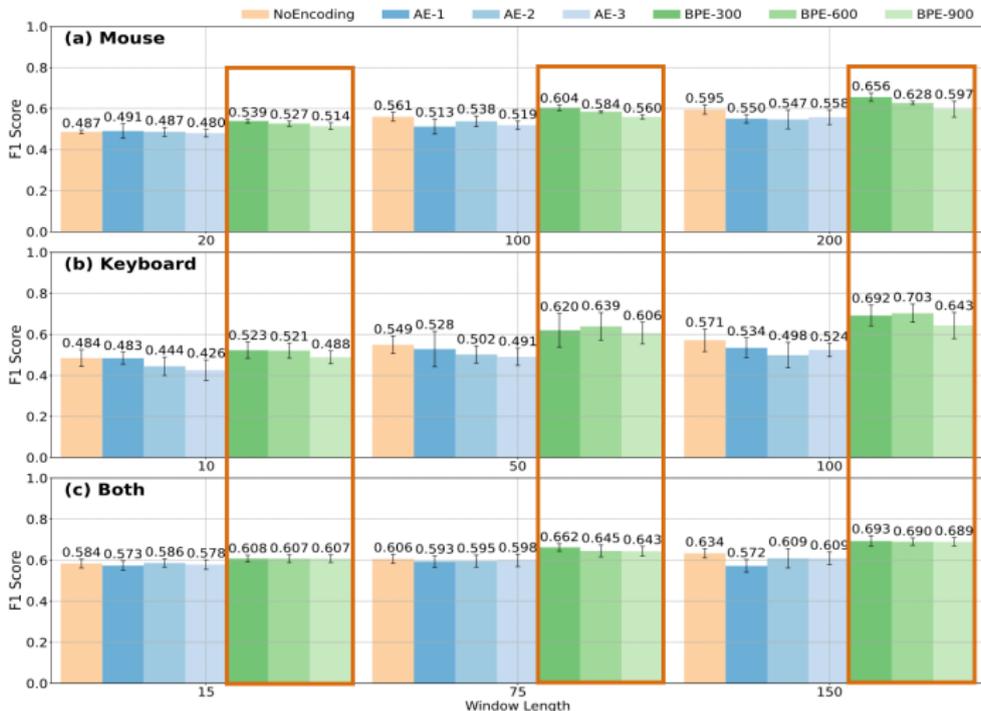
- Interactive behaviour is similar to natural language
- NLP methods can be used to model interactive behaviour



[Zhang INTERACT'23 Best Student Paper Nominees]



## Modelling Interactive Behaviour using NLP Methods



Thank you!



- Hu TVCG'19. Sgaze: a data-driven eye-head coordination model for realtime gaze prediction. *IEEE Transactions on Visualization and Computer Graphics*, 25(5):2002–2010, 2019.
- Hu TVCG'20. Dgaze: Cnn-based gaze prediction in dynamic scenes. *IEEE Transactions on Visualization and Computer Graphics*, 26(5):1902–1911, 2020.
- Hu TVCG'21. Fixationnet: forecasting eye fixations in task-oriented virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, 27(5):2681–2690, 2021.
- Hu TVCG'22. Ehtask: recognizing user tasks from eye and head movements in immersive virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 2022.
- Jiao UIST'23. Supreyes: Super resolution for eyes using implicit neural representation learning. In *Proc. ACM Symposium on User Interface Software and Technology*, pages 1–13, 2023. doi: [10.1145/3586183.3606780](https://doi.org/10.1145/3586183.3606780).
- Zhang INTERACT'23. Exploring natural language processing methods for interactive behaviour modelling. In *Proc. IFIP TC13 Conference on Human-Computer Interaction (INTERACT)*, pages 1–22. Springer, 2023.

