DGaze: CNN-Based Gaze Prediction in Dynamic Scenes

Zhiming Hu[1], Sheng Li[1], Congyi Zhang[2,1], Kangrui Yi[1], Guoping Wang[1], Dinesh Manocha[3]

[1]Peking University

[2]The University of Hong Kong

[3]University of Maryland

Project URL: cranehzm.github.io/DGaze

Outline

- Background
- Related Work
- DGaze Model
- Limitations and Future Work

Eye Tracking Technology

Eye Tracking Technology[1]

[1] https://www.7invensun.com/

Background

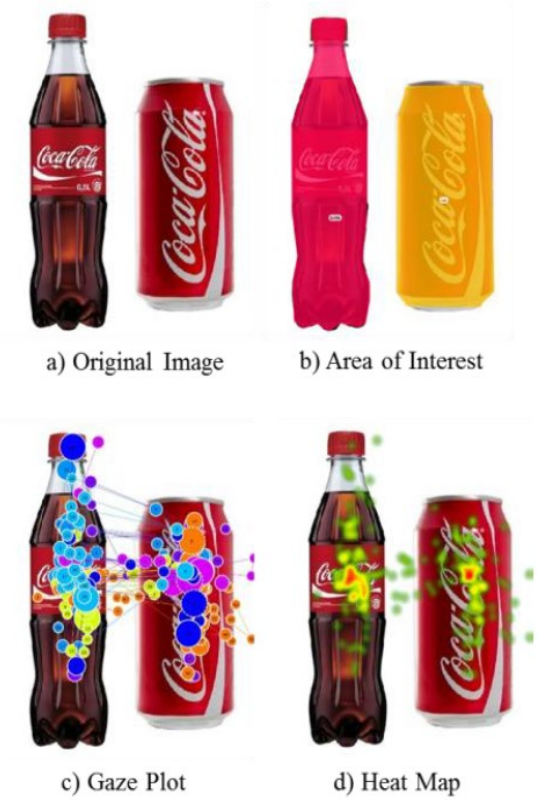
北京大學 PEKING UNIVERSITY

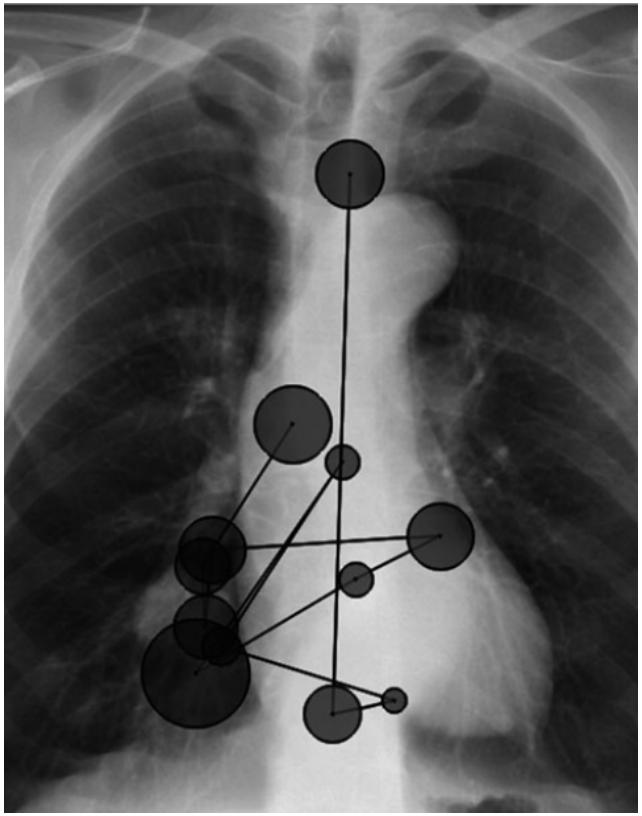
Eye Tracking Technology

- Neuroscience & Psychology
- Industrial Engineering
- Marketing & Advertising
- Computer Science
- .....

北京大學 PEKING UNIVERSITY

Eye Tracking Technology





Marketing Strategy Analysis
[Zamani et al. 2016]

Cognitive Research
[Kiefer et al. 2017]

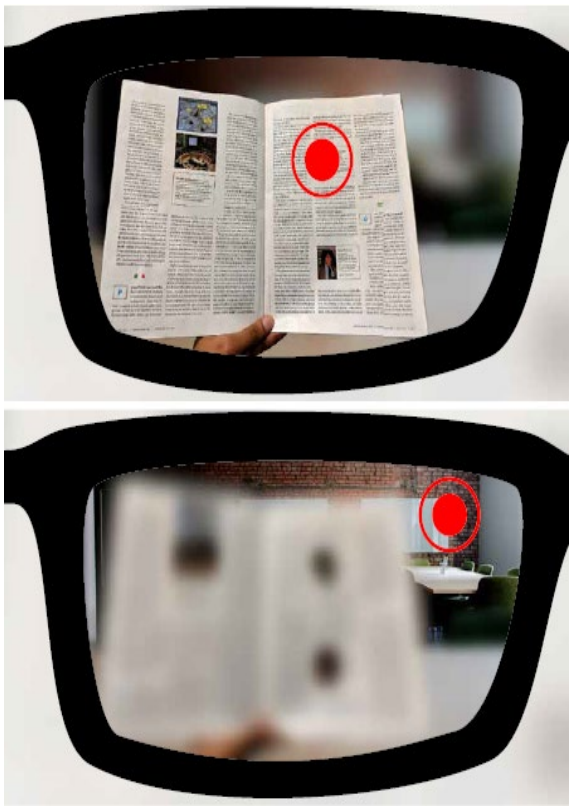
Medical Education
[Kok et al. 2017]

北京大學 PEKING UNIVERSITY

Eye Tracking Technology

Gaze-based Interaction
[Pfeiffer et al. 2008]

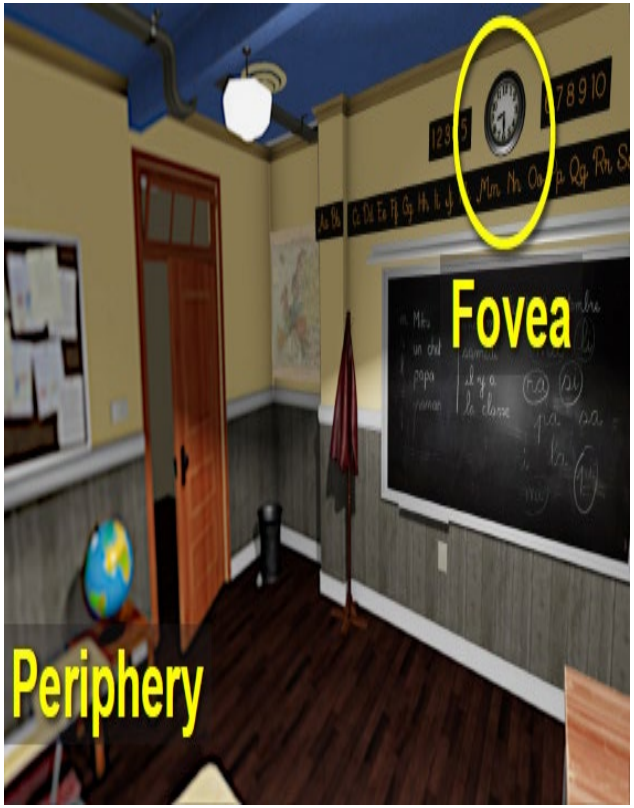
Collaborative System
[Zhang et al. 2017]



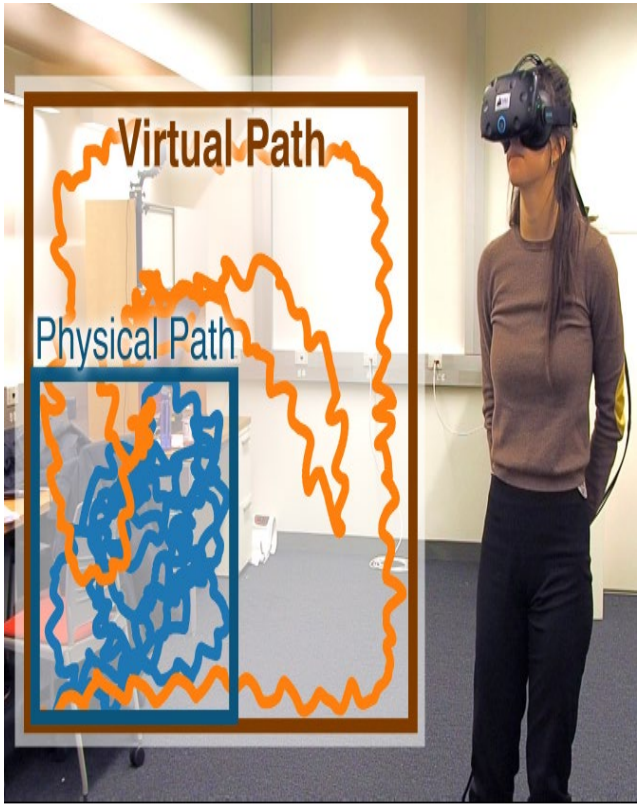
Gaze-contingent Eyeglasses
[Padmanaban et al. 2019]

北京大學 PEKING UNIVERSITY

Eye Tracking in Virtual Reality



Gaze-contingent Rendering
[Patney et al. 2016]



Redirected Walking
[Sun et al. 2018]

Gaze Behavior Analysis
[Alghofaili et al. 2019]

Solution to Eye Tracking in VR

Hardware-based Solution

Eye Tracker[1]

- Accurate ✓
- Currently Expensive ✗
- Not Widely Available ✗
- May Need Calibration ✗
- Cannot Predict Future Gaze Position ✗

[1] https://www.7invensun.com/

北京大学 PEKING UNIVERSITY

Motivation of Our Work

- Propose a **software-based** eye tracking solution in VR that only employs information from the VR system

Our Goals

- Reveal the characteristics of users' gaze behaviors in virtual reality
- Predict users' gaze positions based on the characteristics of users' gaze

北京大學 PEKING UNIVERSITY

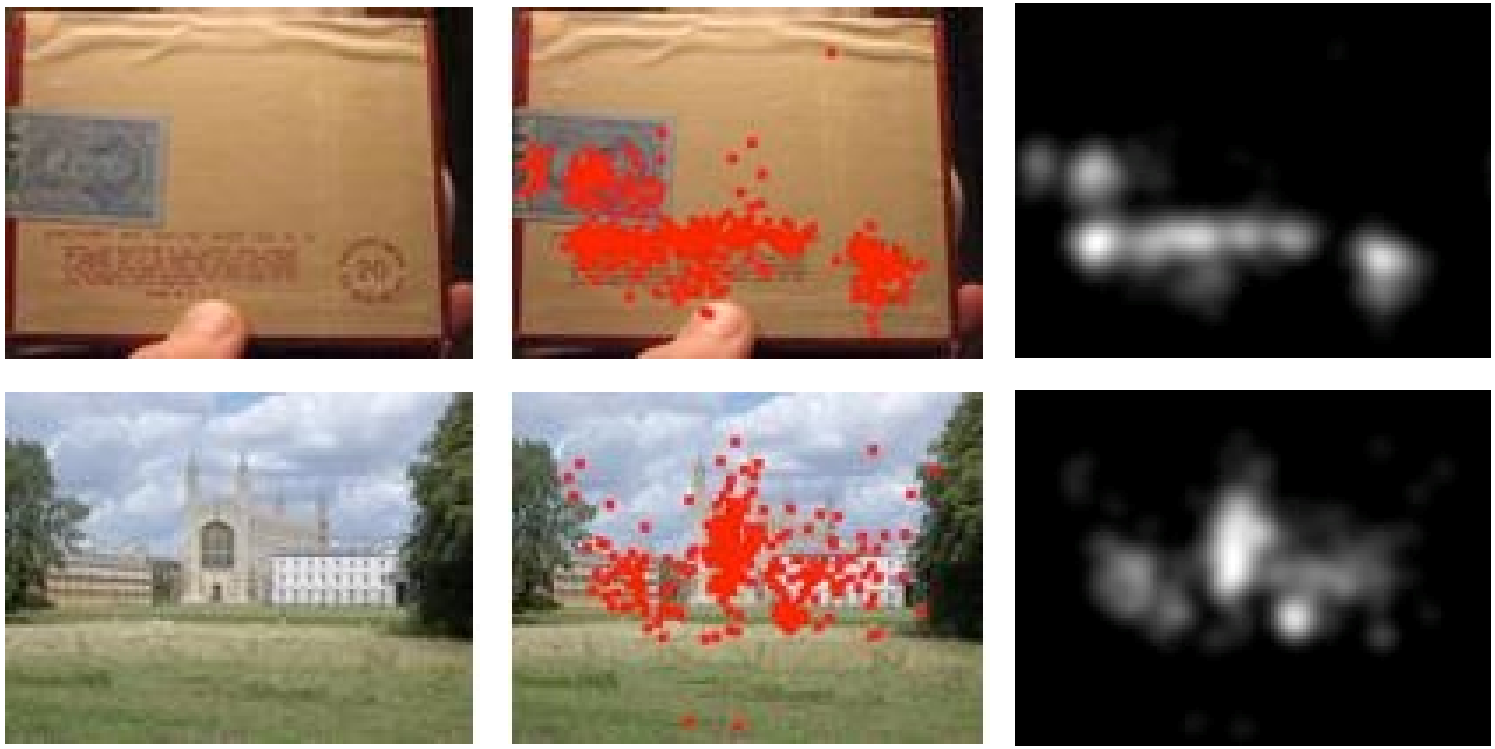
Salient Object Detection

Top: Original Images[1]; Bottom: Salient Objects [1]

[1] https://mmcheng.net/msra10k/

北京大學 PEKING UNIVERSITY

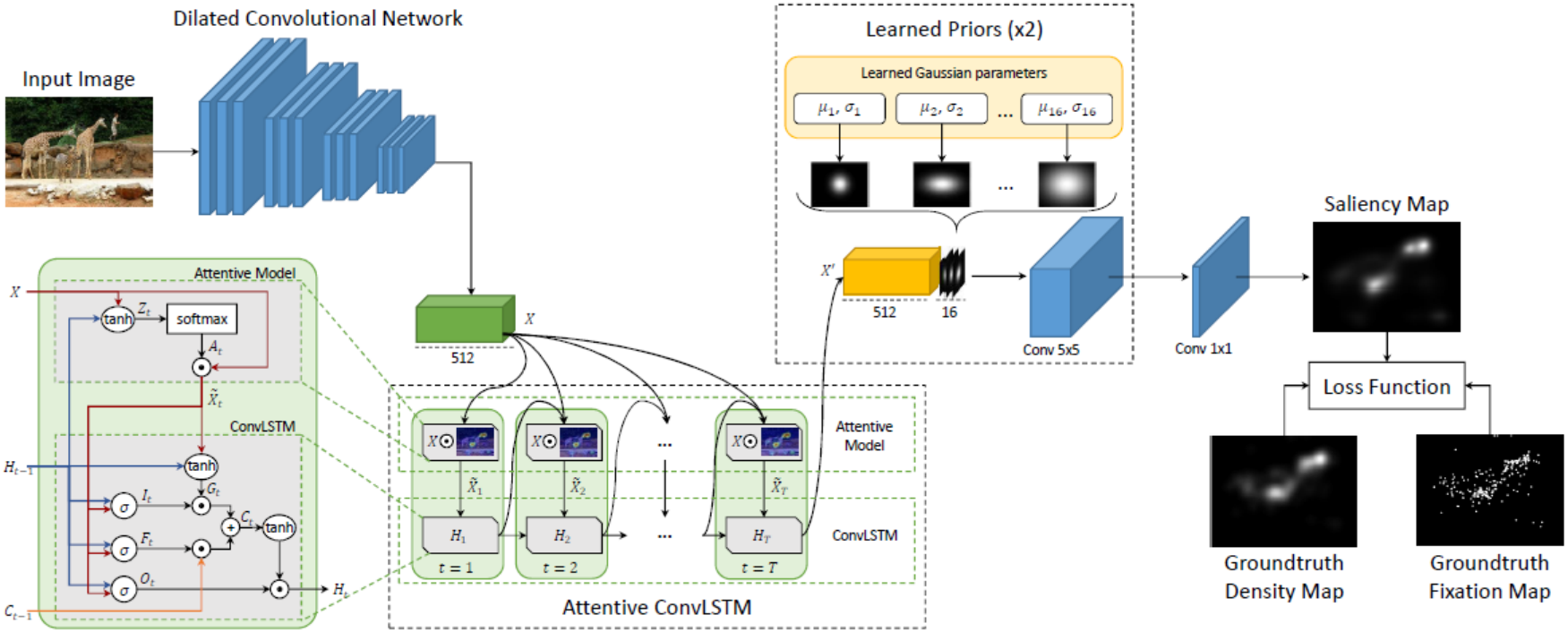
Saliency Prediction



Original Images[1] Eye Fixations[1] Saliency Maps[1]

[1] http://saliency.mit.edu/results_mit300.html

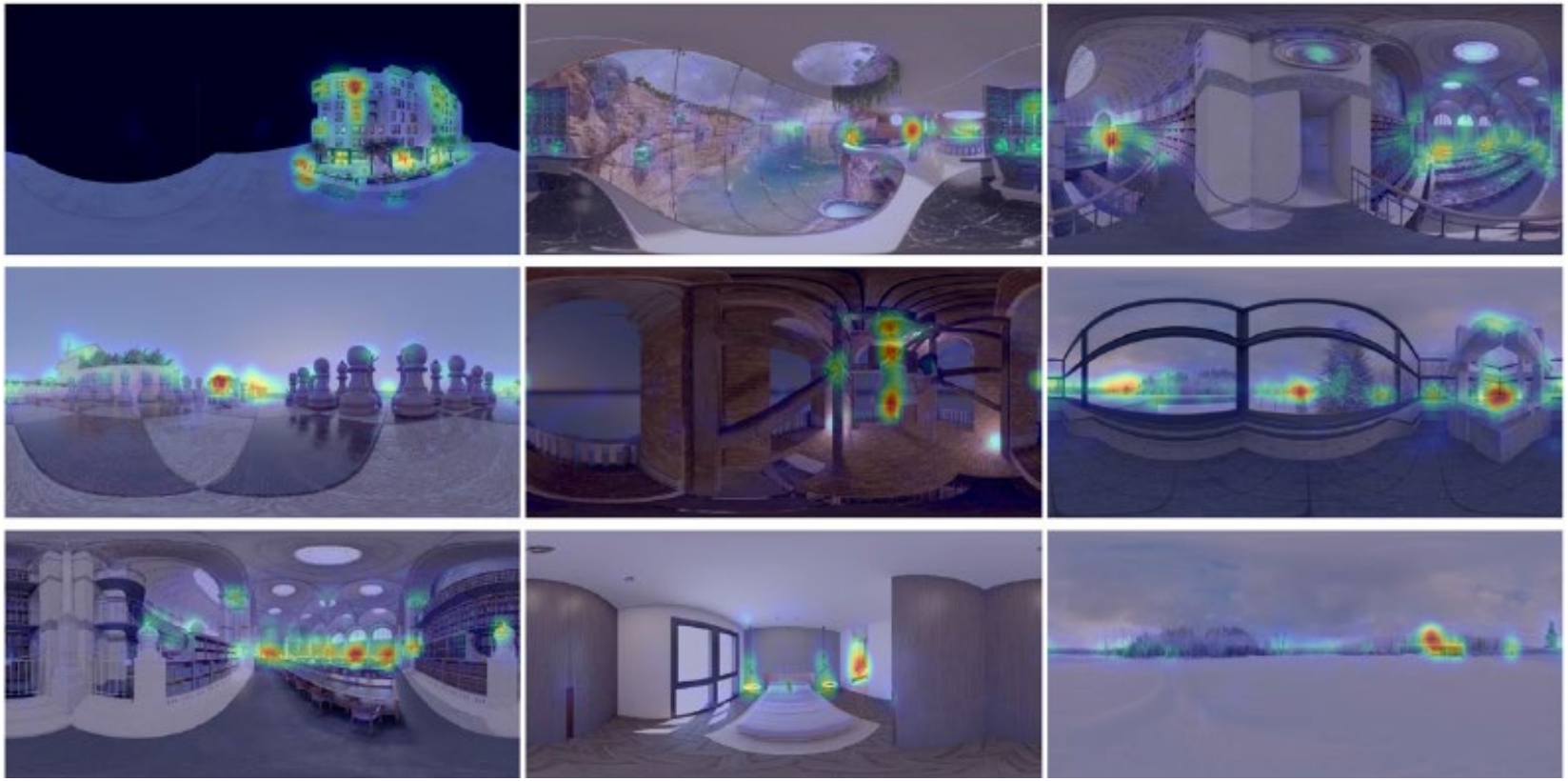
Deep Learning-Based Saliency Predictor



Saliency Attentive Model (SAM)
[Cornia et al. 2018]

北京大學 PEKING UNIVERSITY

Saliency in 360° Images



Saliency in 360° Images
[Sitzmann et al. 2018]

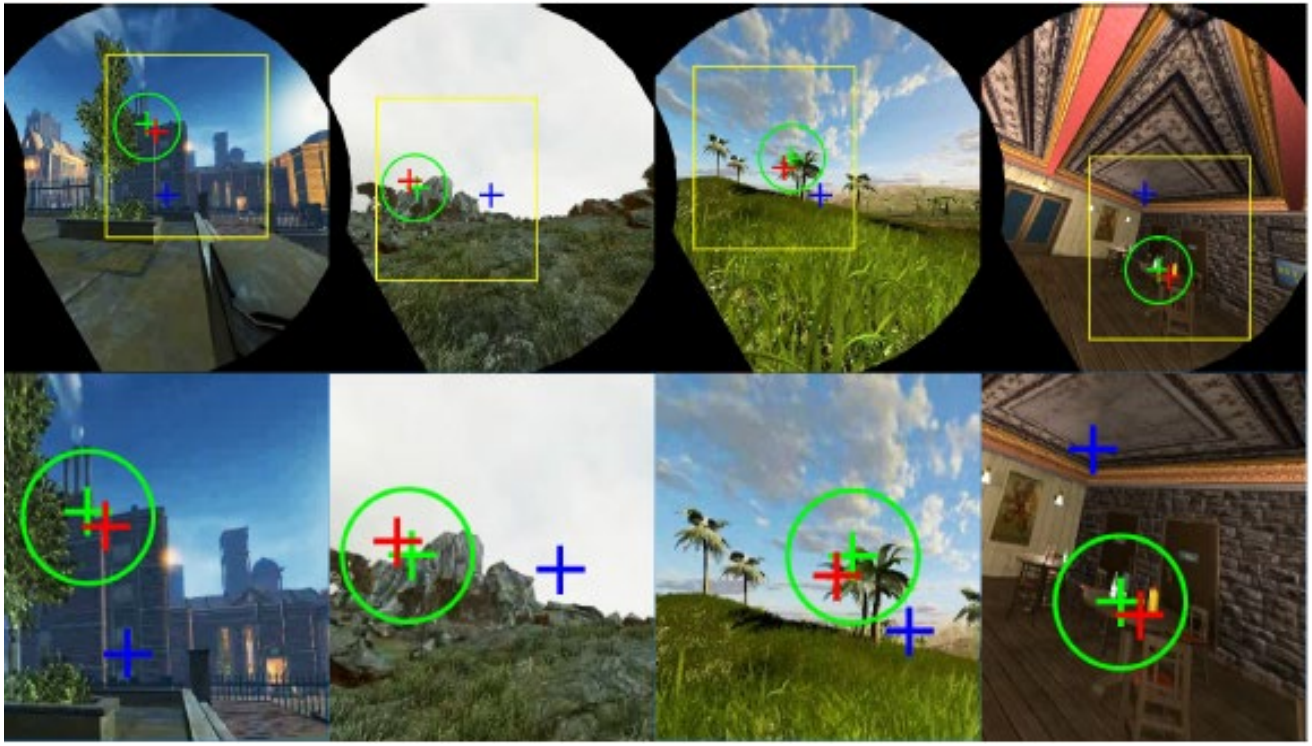
北京大學 PEKING UNIVERSITY

Saliency in 360° Videos

Saliency in 360° Videos
[Xu et al. 2018]

北京大學 PEKING UNIVERSITY

Gaze Prediction in Static Virtual Scenes



Gaze Prediction in Static Virtual Scenes
[Hu et al. 2019]

北京大學 PEKING UNIVERSITY

Our Work *vs.* Previous Work

- Goal: **2D gaze positions** *vs.* salient objects/saliency maps
- Scene: **3D virtual scenes** *vs.* images/videos

  dynamic scenes *vs.* static scenes

Challenges

- Gaze position prediction in VR requires higher accuracy than saliency prediction
- Gaze behavior in 3D scenes are different from that in 2D scenes
- Dynamic scenes are more intricate than static scenes

北京大學 PEKING UNIVERSITY

Contributions

- Propose a novel CNN-based gaze prediction model (DGaze)
- Provide comprehensive analyses of human gaze behaviors in dynamic virtual scenes
- Build an eye tracking dataset that contains 43 users' gaze data in 5 dynamic scenes

Workflow

- Data Collection
- Gaze Behavior Analysis
- CNN-Based Gaze Prediction Model (DGaze)
- Model Evaluation

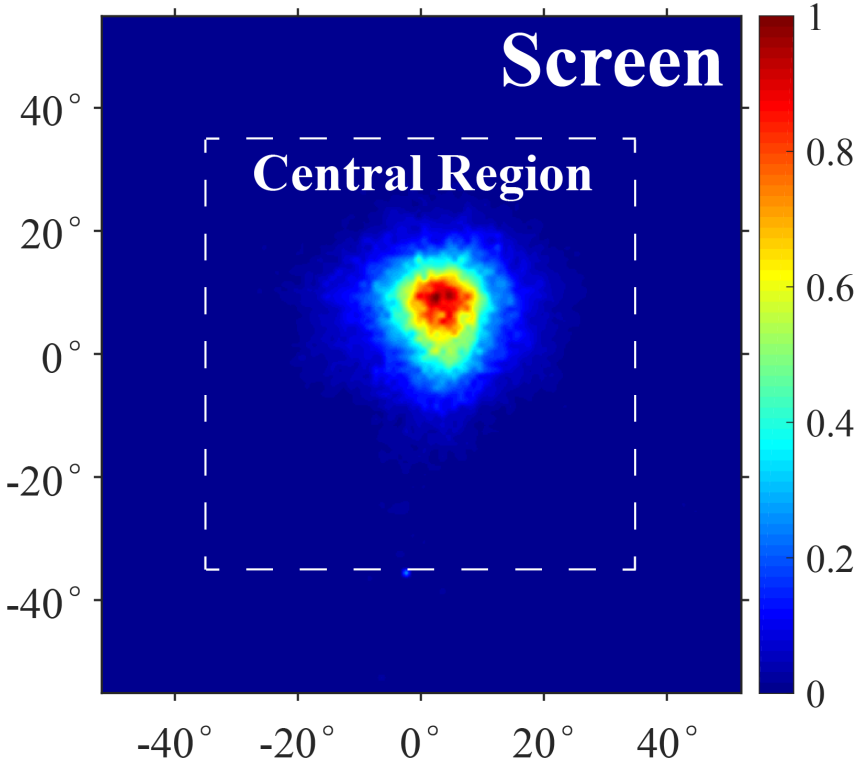
北京大學 PEKING UNIVERSITY

Data Collection

- Participants: 43 users (25 male, 18 female, ages 18-32)
- Stimuli: 5 dynamic virtual scenes
- System: HTC Vive + eye tracker
- Procedure: free-viewing, no task
- Data: scene screenshots + gaze positions + head poses + dynamic object positions

Stimuli

Gaze Behavior Analysis: Gaze Analysis

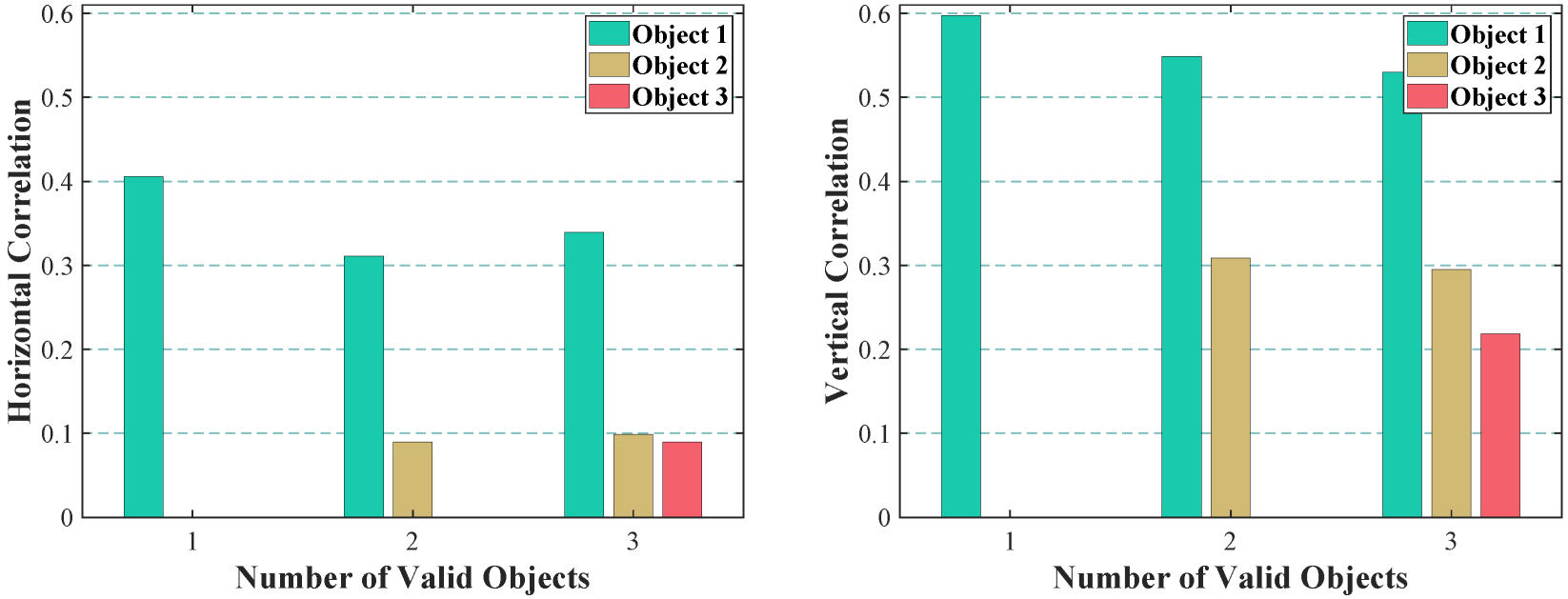


The distribution of users' gaze positions on the HMD's screen

Most of the gaze data lies in the central region of the screen.

Gaze Behavior Analysis: Gaze-Object Analysis

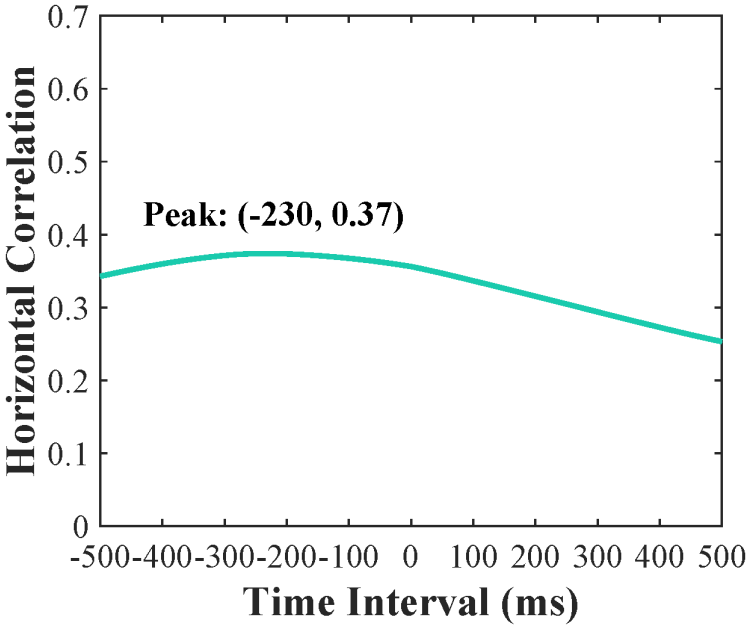
Spearman's rank correlation coefficient

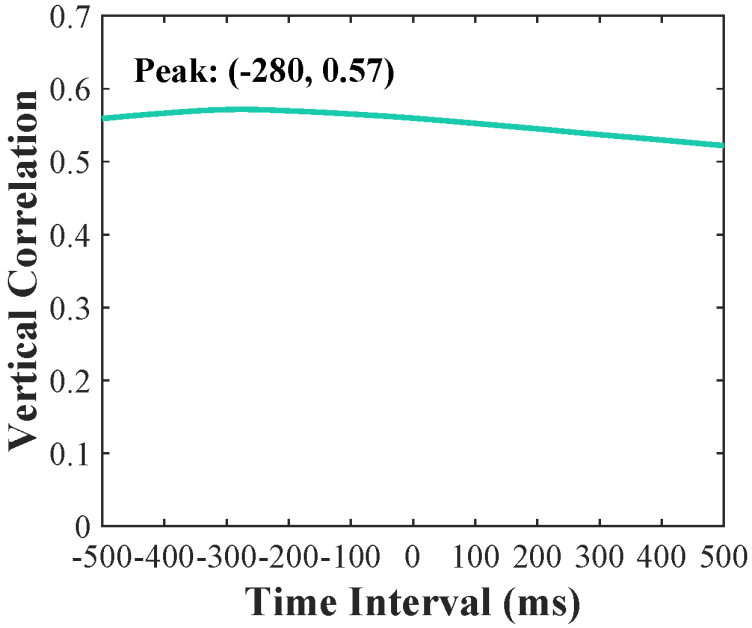


The horizontal (left) and vertical (right) correlations between gaze positions and object positions

Users' gaze positions are strongly correlated with dynamic object positions.

Gaze Behavior Analysis: Gaze-Object Analysis

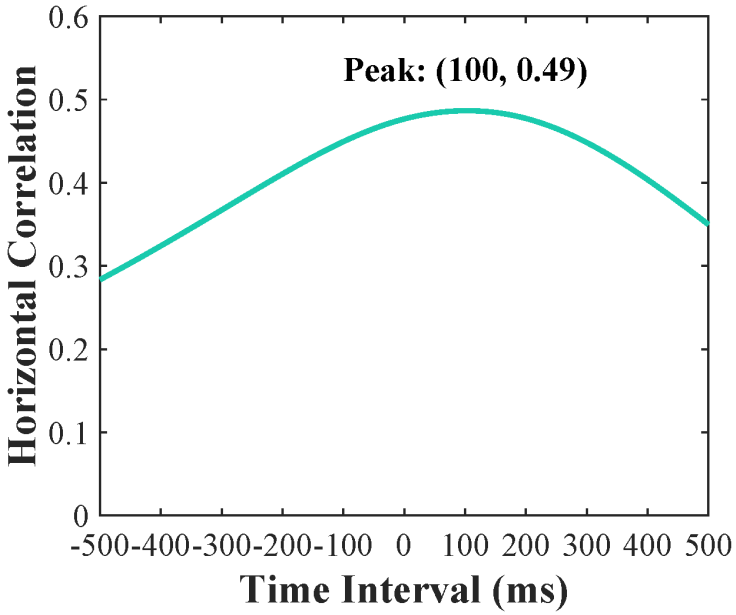


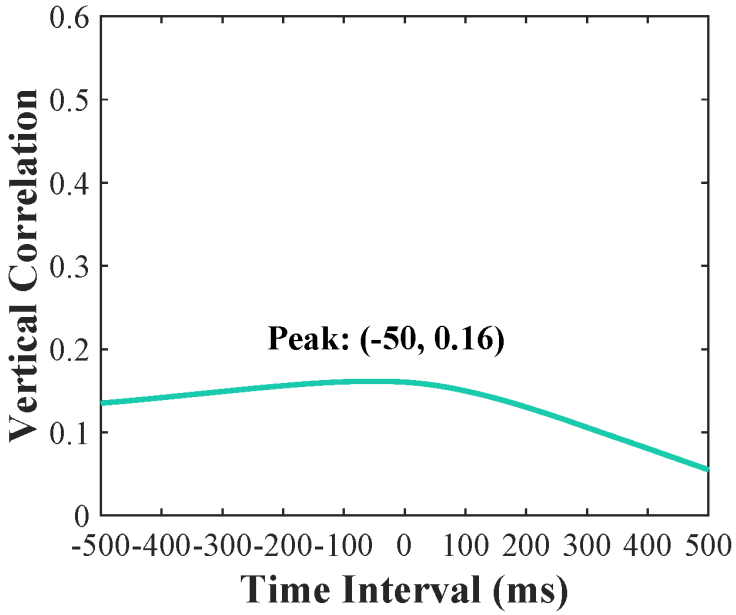


The correlations between gaze positions and the nearest object positions at different time intervals

Both realtime and past object positions are correlated with gaze positions.

Gaze Behavior Analysis: Gaze-Head Analysis



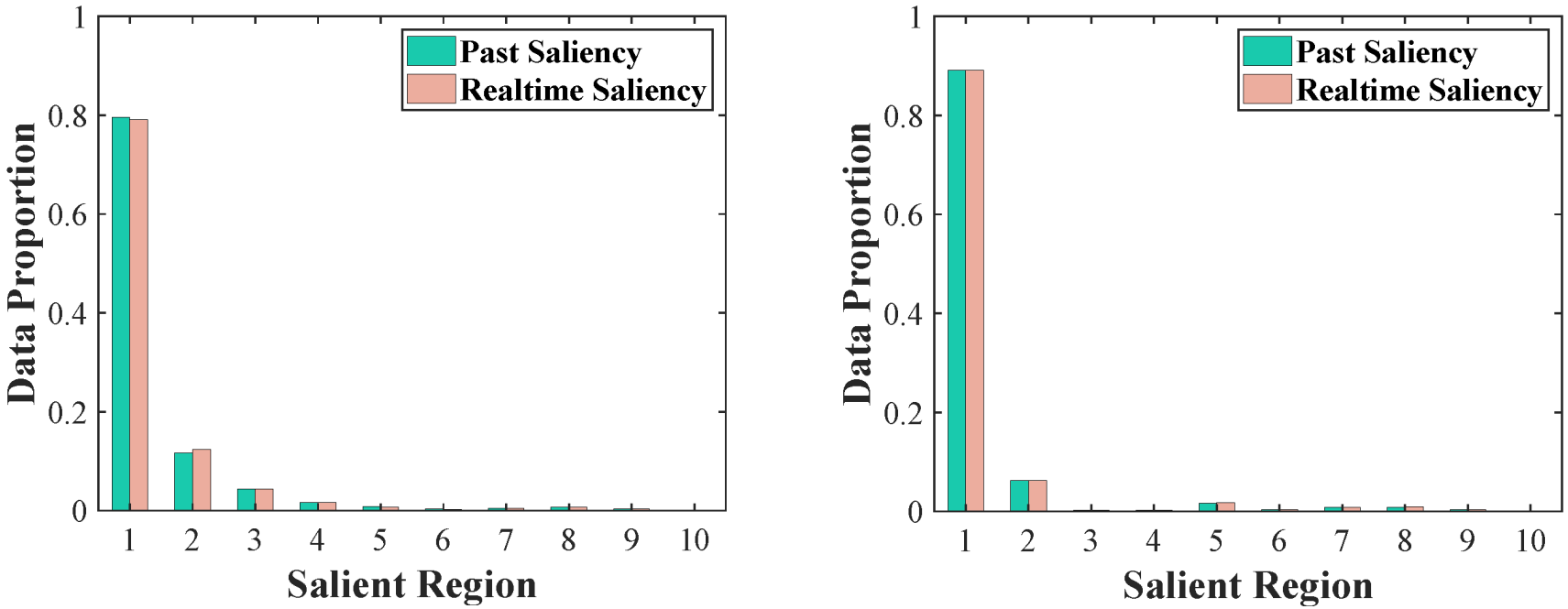


The correlations between gaze positions and head velocities at different time intervals

Both realtime and past head velocities are correlated with gaze positions.

北京大學 PEKING UNIVERSITY

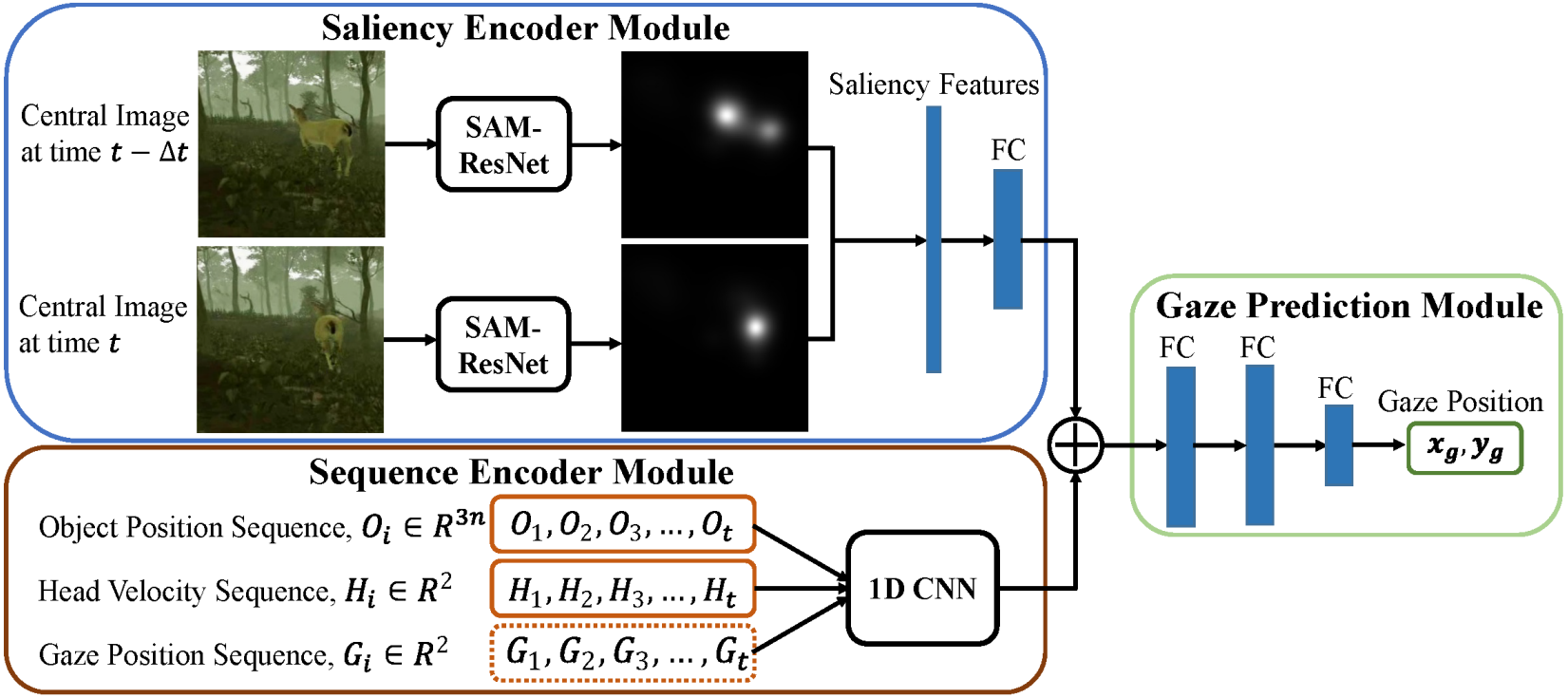
Gaze Behavior Analysis: Gaze-Saliency Analysis



The distributions of gaze positions on salient regions of the whole image (left) and the central image (right)

Most of the gaze positions lie in the most salient region (region 1).

CNN-Based Gaze Prediction Model (DGaze)



Architecture of DGaze model

DGaze_ET: predict future gaze positions with higher precision by combining accurate past gaze data.

CNN-Based Gaze Prediction Model (DGaze)

- Saliency Encoder Module: extract and encode the saliency features of the VR content
- Sequence Encoder Module: encode the dynamic object position sequence, the head velocity sequence, and the gaze position sequence (DGaze_ET).
- Gaze Prediction Module: combine the outputs of the above 2 modules to predict users' gaze positions.

北京大學 PEKING UNIVERSITY

Model Evaluation: Realtime Prediction

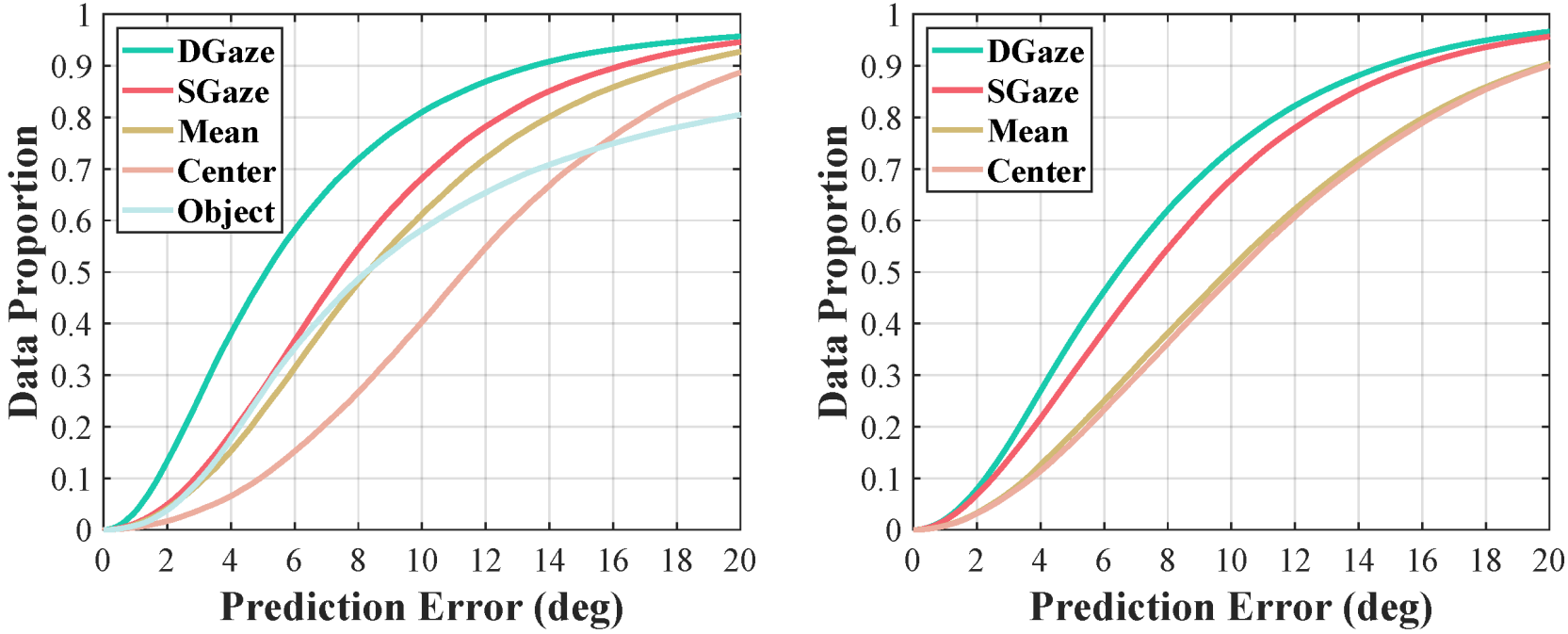
		DGaze	SGaze	Mean	Center	Object
Dynamic	Mean	7.11°	9.11°	10.04°	12.46°	13.25°
	SEM	0.01°	0.01°	0.01°	0.01°	0.02°
Static	Mean	7.71°	8.52°	10.93°	11.16°	
	SEM	0.01°	0.01°	0.01°	0.01°	

DGaze and other methods' realtime prediction performances on the dynamic dataset and the static dataset

DGaze performs best in both dynamic and static scenes.

北京大學 PEKING UNIVERSITY

Model Evaluation: Realtime Prediction

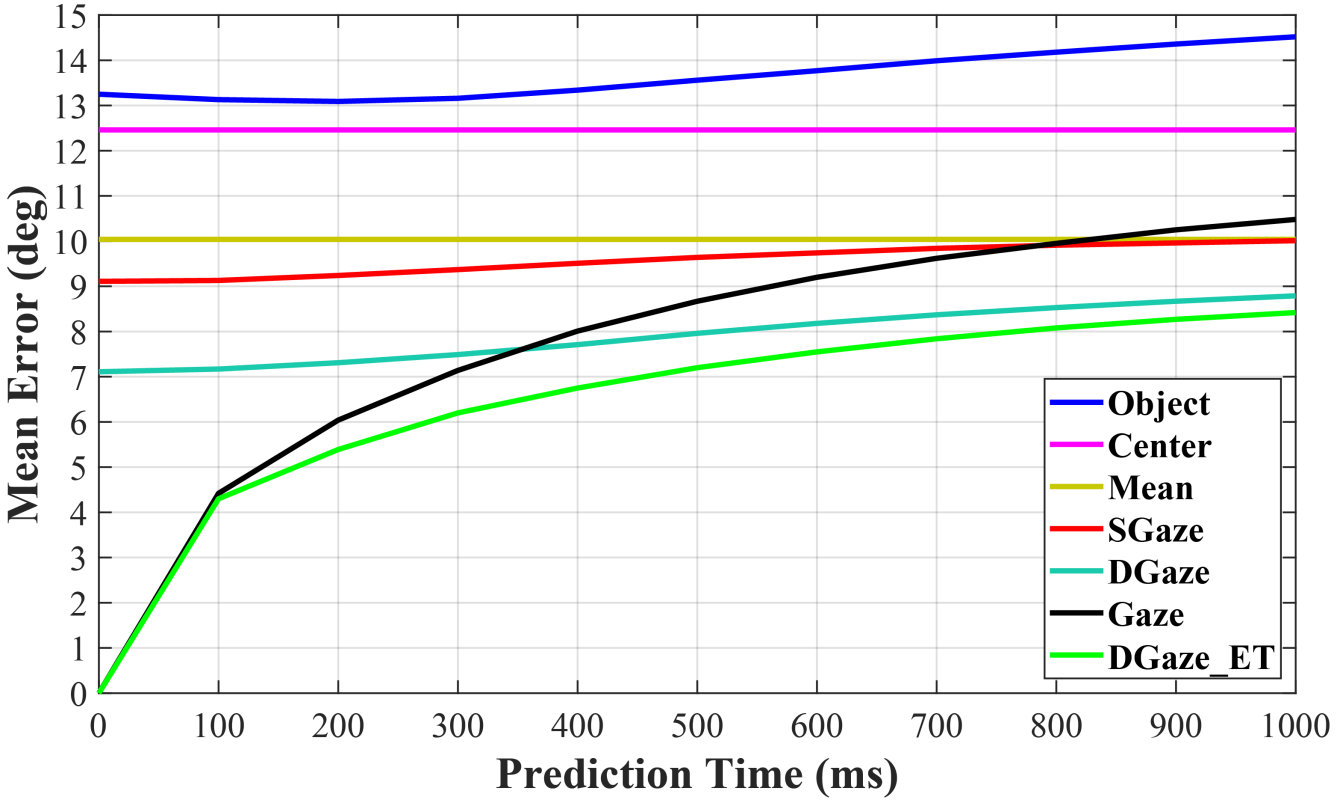


Cumulative distribution function (CDF) of the prediction errors on the dynamic dataset (left) and the static dataset (right)

DGaze performs best in terms of CDF curve.

北京大學 PEKING UNIVERSITY

Model Evaluation: Future Prediction



DGaze and other methods' future prediction performances in dynamic scenes

DGaze and DGaze_ET outperform other methods in different prediction times.

Gaze-Contingent Rendering

Gaze-Contingent Rendering

User Study

DGaze *vs.* prior method

t-test, $p < 0.01$

DGaze performs significantly better than prior method.

北京大學 PEKING UNIVERSITY

Task-Oriented Game

Game Scene

Limitations

- Our dataset is restricted to free-viewing conditions
- The type of dynamic objects used in the experiments is limited
- The influence of sound is not considered in our model

Future Work

- Overcome the limitations
- Improve our model's performance by fine-tuning the parameters
- Extend our model to consider more input features
- Convert our model to other systems like AR and MR systems

Thank you!