

虚拟现实环境中用户视觉注意 的分析与预测

胡志明 1701111311

信息科学技术学院

计算机软件与理论专业

研究方向：人机交互与虚拟现实

导师：汪国平教授

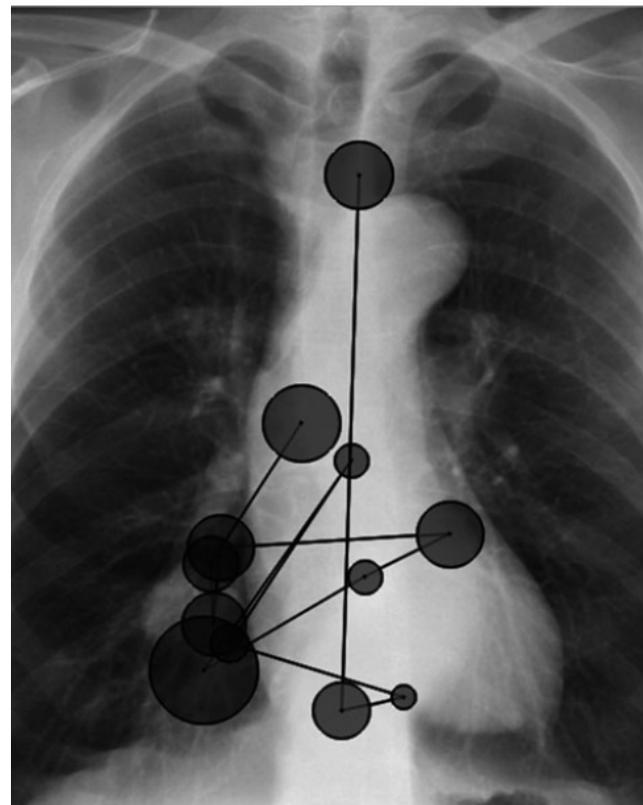
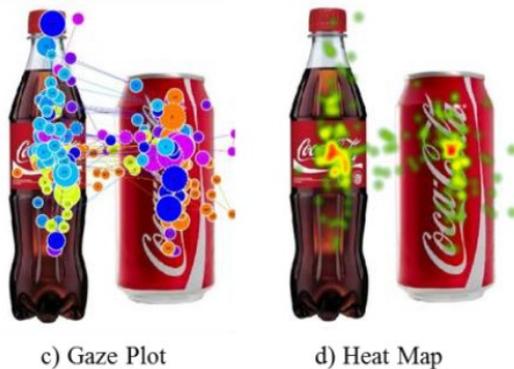
- ▶ 研究背景
- ▶ 相关工作
- ▶ 基于眼动头动协调性的静态虚拟场景注视预测模型
- ▶ 基于卷积神经网络的动态虚拟场景注视预测模型
- ▶ 基于学习的任务驱动虚拟场景注视预测模型
- ▶ 基于学习的虚拟现实环境用户任务识别模型
- ▶ 总结与展望

用户视觉注意



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

视觉注意的应用



营销策略分析
[Zamani et al. 2016]

感知研究
[Kiefer et al. 2017]

医学教育
[Kok et al. 2017]

视觉注意的应用



眼动交互
[Pfeiffer et al. 2008]

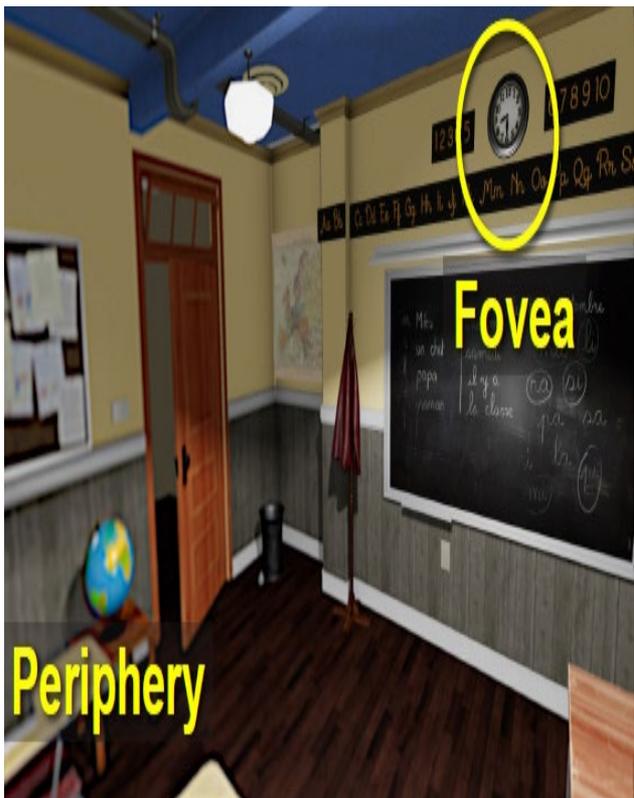


协作系统
[Zhang et al. 2017]

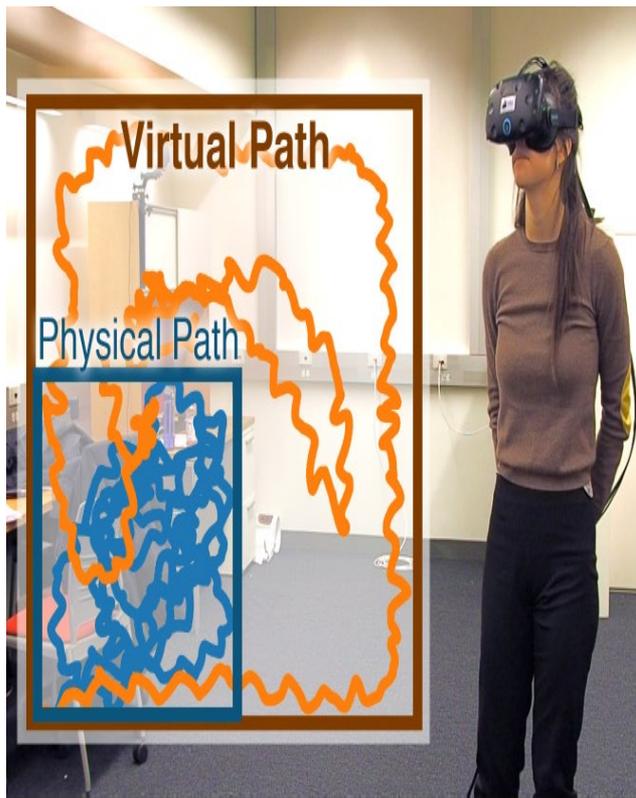


注视跟随眼镜
[Padmanaban et al. 2019]

视觉注意在虚拟现实中的应用



注视点渲染
[Patney et al. 2016]



注视引导
[Sun et al. 2018]



场景布局优化
[Alghofaili et al. 2019]

视觉注意在虚拟现实中的应用



虚拟现实内容生成
[Sitzmann et al. 2018]

细节层次管理
[Lee et al. 2008]

虚拟现实系统智能化
[Hadnett-Hunter et al. 2019]

视觉注意与数字人形象生成



[<https://www.vive.com/us/product/vive-pro-eye/features/>]

眼动仪的缺陷

▶ 价格昂贵

- Pupil VR/AR (VR/AR眼动仪) 1750欧元
- Pupil Invisible (眼镜式眼动仪) 5500欧元

▶ 不能预测用户在未来时刻的注视

- 眼动仪只能提供用户在当前时刻的注视点

▶ 不能揭示用户视觉注意的深层信息

- 用户正在执行的任务、用户的心理负荷、用户的交互意图

显著性物体检测



上：原始图像；下：显著性物体

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

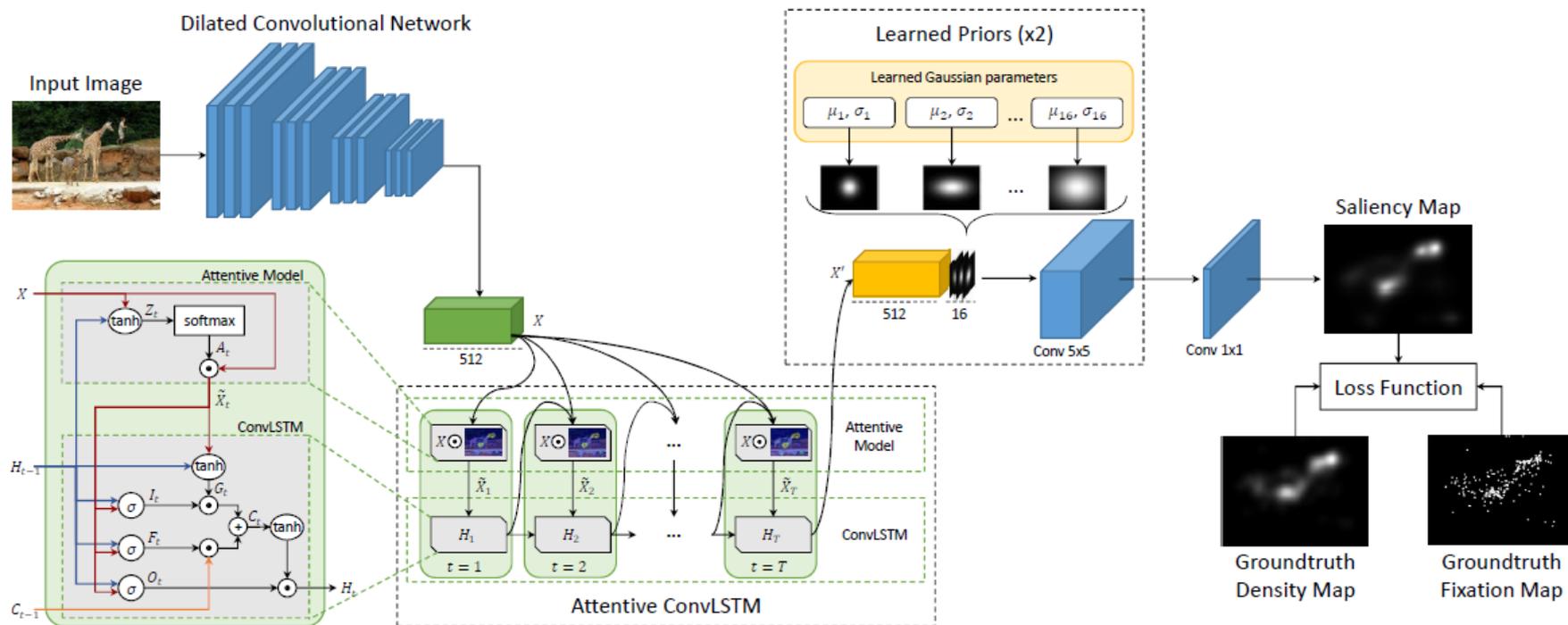
显著性图预测



上：原始图像；下：显著性图

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

基于学习的显著性预测模型



SAM模型
[Cornia et al. 2018]

二维图片场景用户任务识别



Two goldfish, named Shaggy and Daphne, have become the smallest and hardest survivors of the devastating February earthquake in Christchurch, New Zealand. The fish spent four and a half months trapped in their tank in the city's off-limits downtown without anyone to feed them or even any electricity to power their tank filter.

➤ 场景记忆

➤ 阅读

➤ 场景搜索

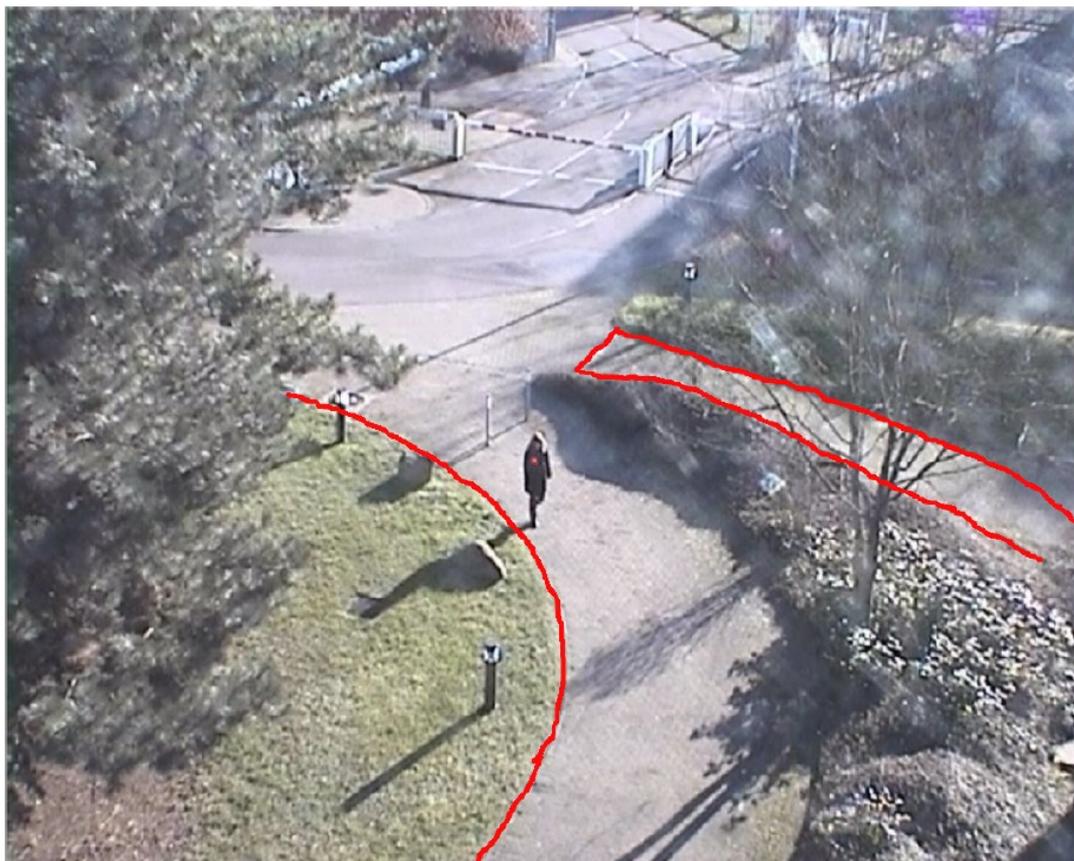
➤ 伪阅读



The goldfish, named Shaggy and Daphne, have become the smallest and hardest survivors of the devastating February earthquake in Christchurch, New Zealand. The fish spent four and a half months trapped in their tank in the city's off-limits downtown without anyone to feed them or even any electricity to power their tank filter.

二维图片场景
[Henderson et al. 2013]

二维视频场景用户任务识别



- 探索
- 观察
- 搜索
- 追踪

二维视频场景
[Hild et al. 2018]

本文工作 vs. 相关工作

- ▶ 研究目标： 注视预测 vs. 显著性物体检测、显著性图预测
- ▶ 实验场景： 虚拟现实环境 vs. 二维图像或视频
- ▶ 输入特征： 场景内容、用户动作信息、物体信息 vs. 场景内容

虚拟现实环境中用户视觉注意分析与预测的优势

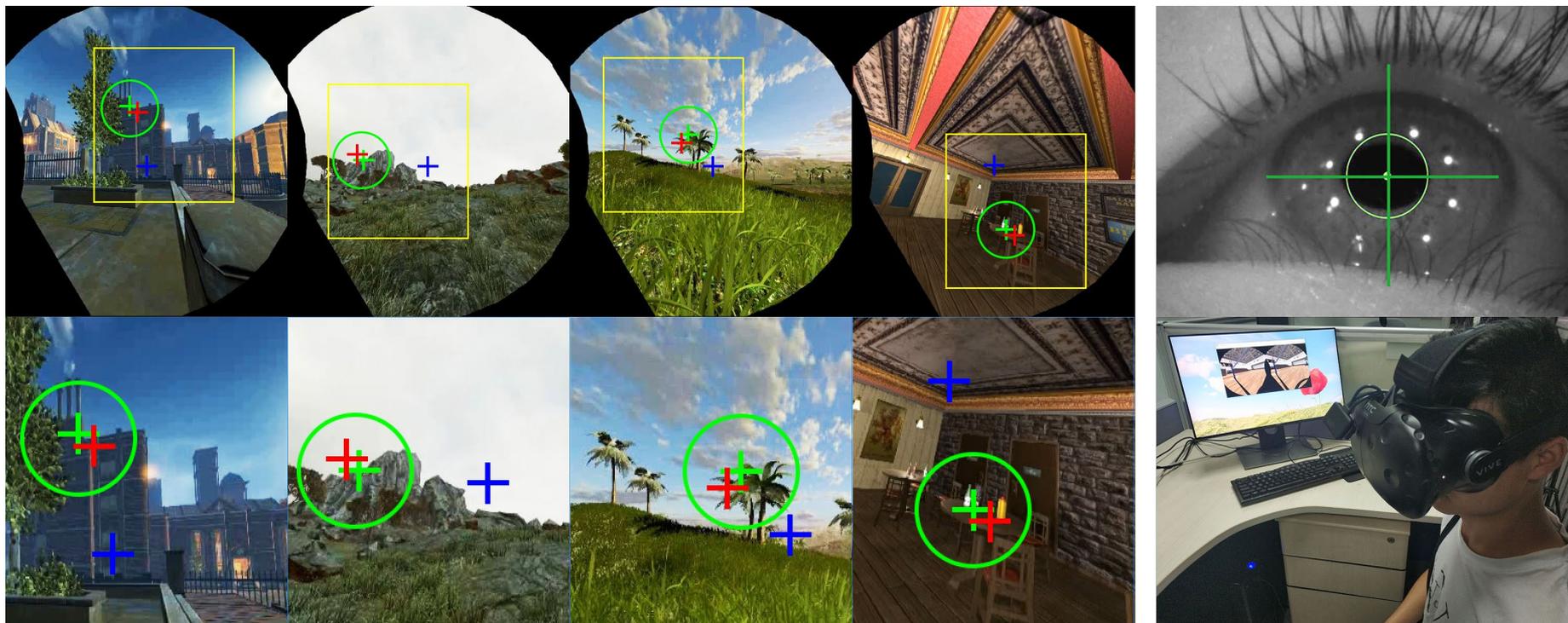
▶ 可以直接获取虚拟场景的全部信息

- 场景内容信息
- 三维物体信息：位置、大小、距离等
- 事件信息：运动物体信息、任务相关物体信息等

▶ 可以直接获取用户动作信息

- 眼睛运动信息
- 头部运动信息
- 手势动作信息

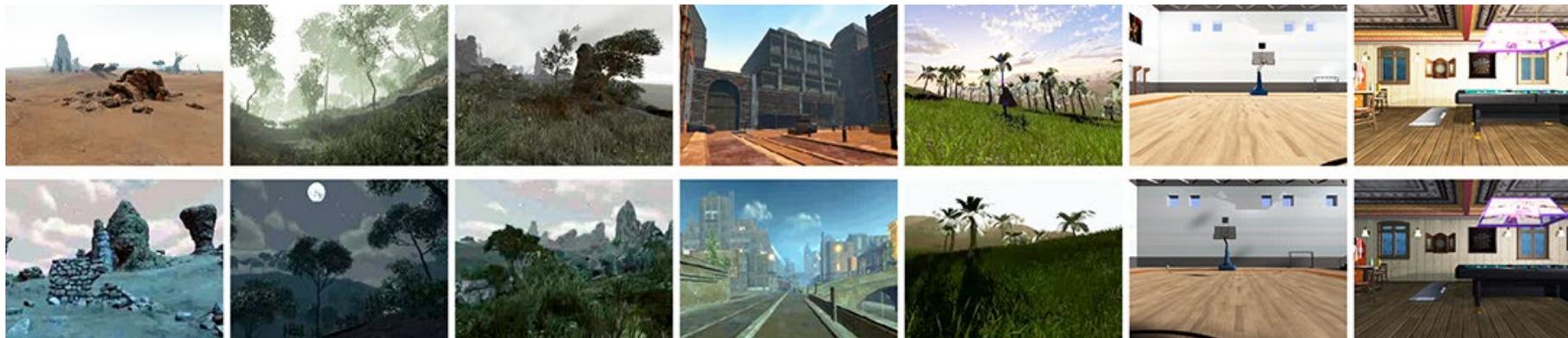
静态虚拟场景注视预测



[Hu et al. TVCG 2019]

注视数据收集

- ▶ 参与者：60名用户（35男，25女，年龄18-36）
- ▶ 实验场景：七个静态虚拟场景
- ▶ 实验设备：HTC Vive头盔、眼动仪
- ▶ 实验过程：自由观察
- ▶ 数据：场景内容、眼睛运动、头部运动



实验场景

眼动头动线性相关性

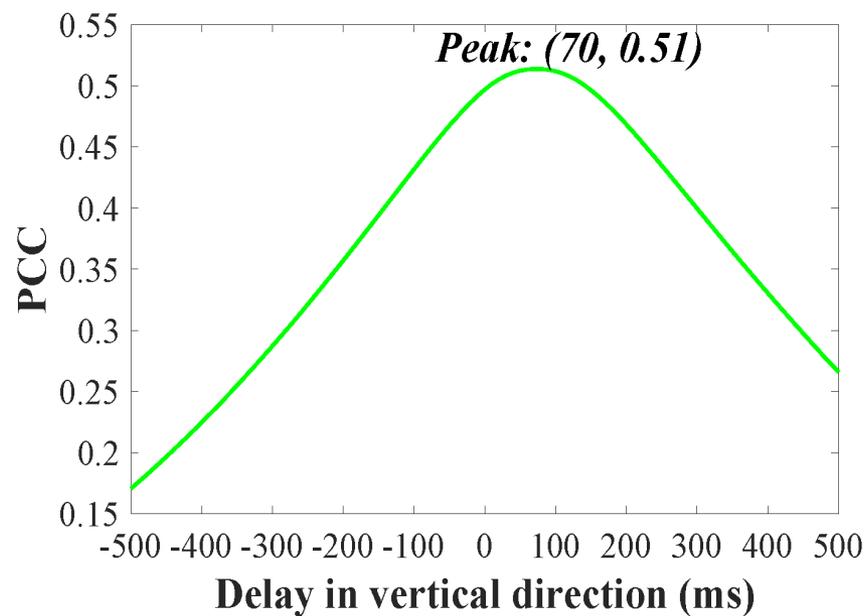
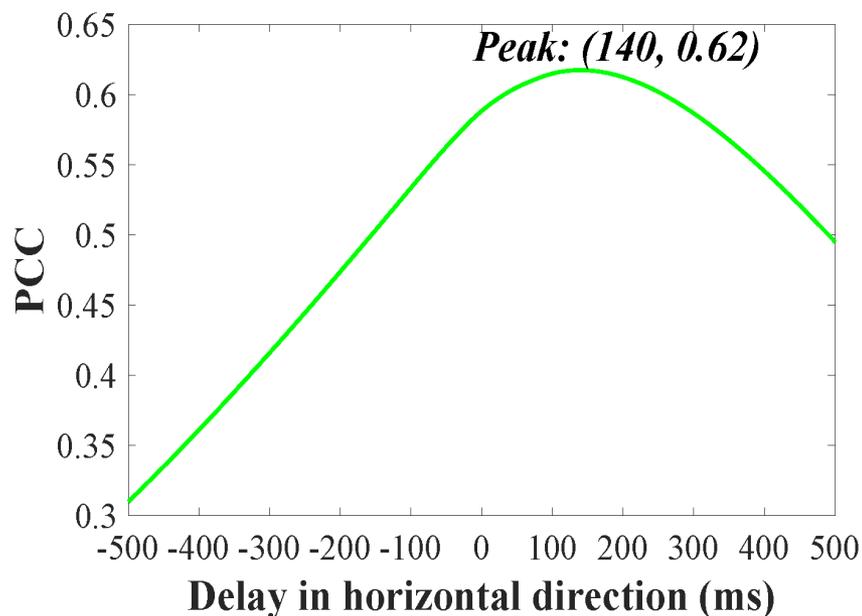
皮尔逊相关因子 (Pearson's correlation coefficient, PCC)

	Static	Intentional	Sudden	Whole
$PCC(v_x)$	0.1345	0.5883	0.1511	0.5641
$PCC(v_y)$	0.1484	0.4969	-0.0906	0.4132

用户注视位置与头部运动之间的相关性

注视位置与头动速度具有很强的线性相关性

眼动头动延迟



注视位置与不同时刻头动速度之间的相关性
左：水平方向；右：竖直方向

头动落后于眼动

SGaze模型

$$x_g(t) = \alpha_x \cdot v_{hx}(t + \Delta t_{x1}) + \beta_x \cdot a_{hx}(t) + F_x(t + \Delta t_{x2}) + G_x(t) + H_x(t)$$

$$y_g(t) = \alpha_y \cdot v_{hy}(t + \Delta t_{y1}) + F_y(t + \Delta t_{y2}) + G_y(t) + H_y(t)$$

x_g, y_g : 注视位置

v_{hx}, v_{hy}, a_{hx} : 头动速度及加速度

F_x, F_y : 内容

G_x, G_y : 任务

H_x, H_y : 其他因素

$\alpha_x, \alpha_y, \beta_x$: 头动速度及加速度的线性影响系数

$\Delta t_{x1}, \Delta t_{y1}$: 眼动头动延迟

眼动头动线性相关性

眼动头动延迟

模型评估

其他方法: Center, Mean, Saliency

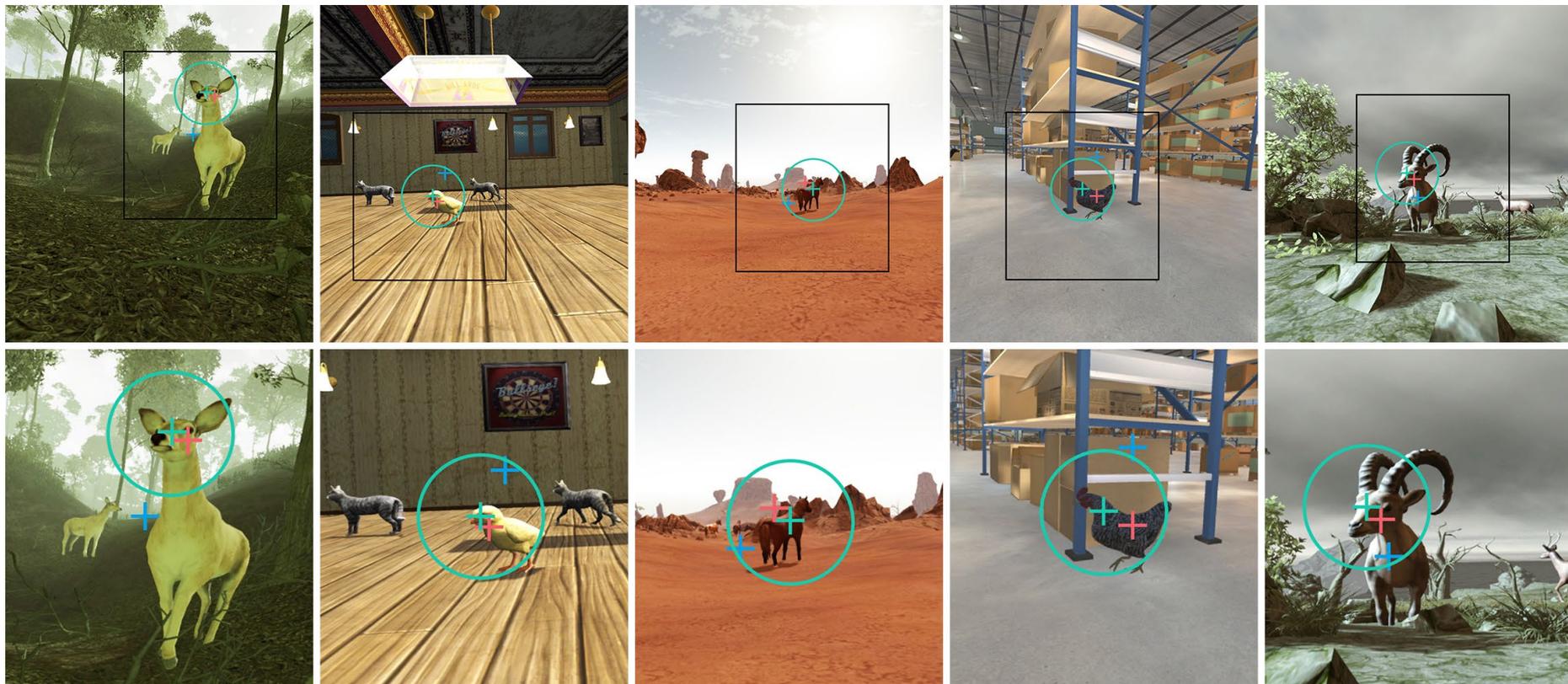
评价指标: 视角误差

	SGaze	Mean	Center	Saliency
Mean Error	8.52°	10.93°	11.16°	21.23°

SGaze模型与其他方法的预测误差均值

SGaze模型的预测效果具有显著的提升

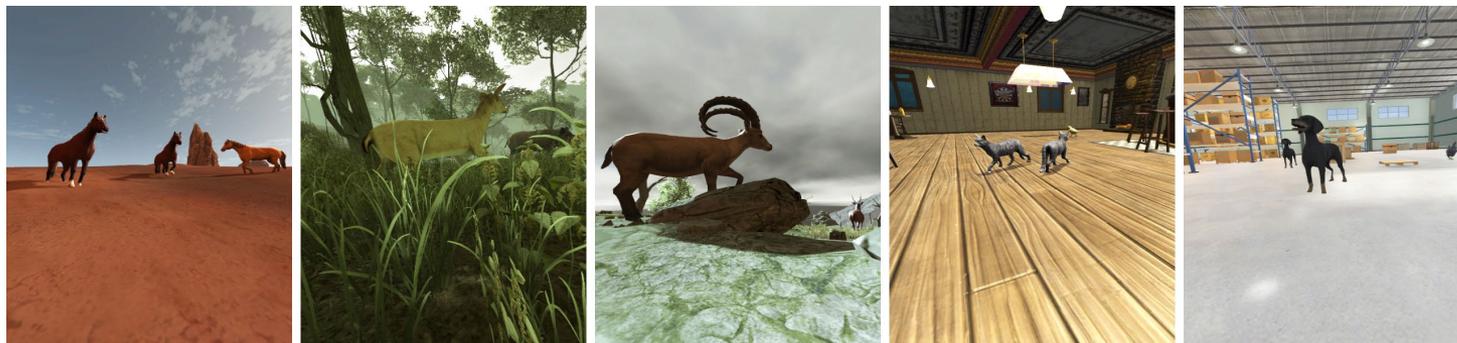
动态虚拟场景注视预测



[Hu et al. TVCG 2020]

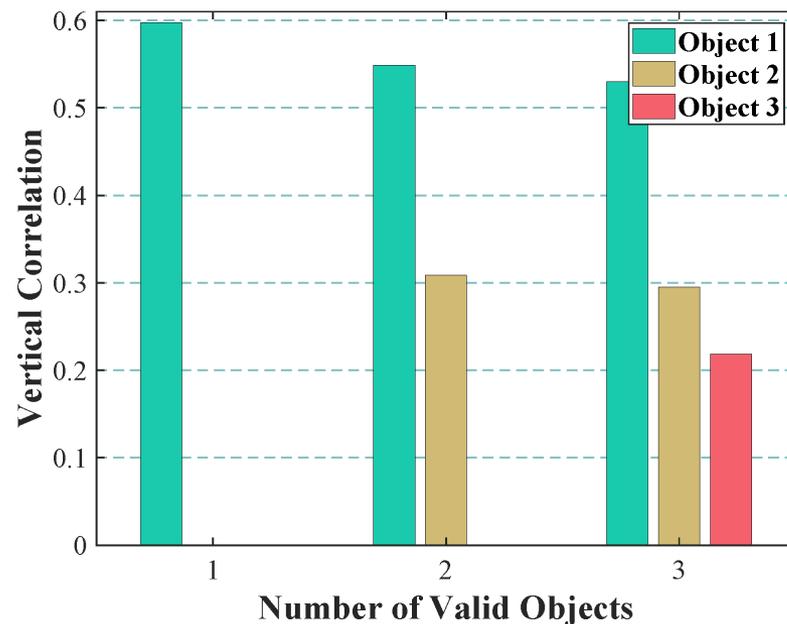
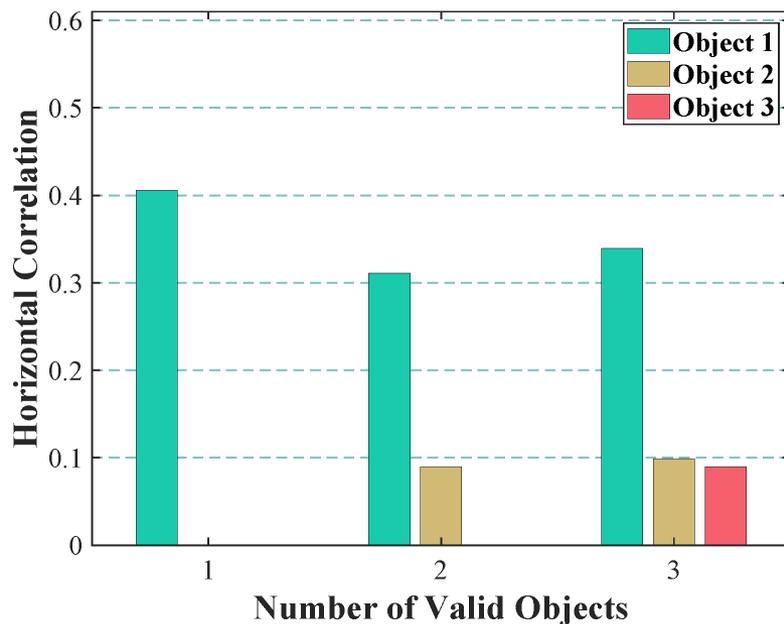
注视数据收集

- ▶ 参与者：43名用户（25男，18女，年龄18-32）
- ▶ 实验场景：五个动态虚拟场景
- ▶ 实验设备：HTC Vive头盔、眼动仪
- ▶ 实验过程：自由观察
- ▶ 数据：场景内容、动态物体位置、眼睛运动、头部运动



实验场景

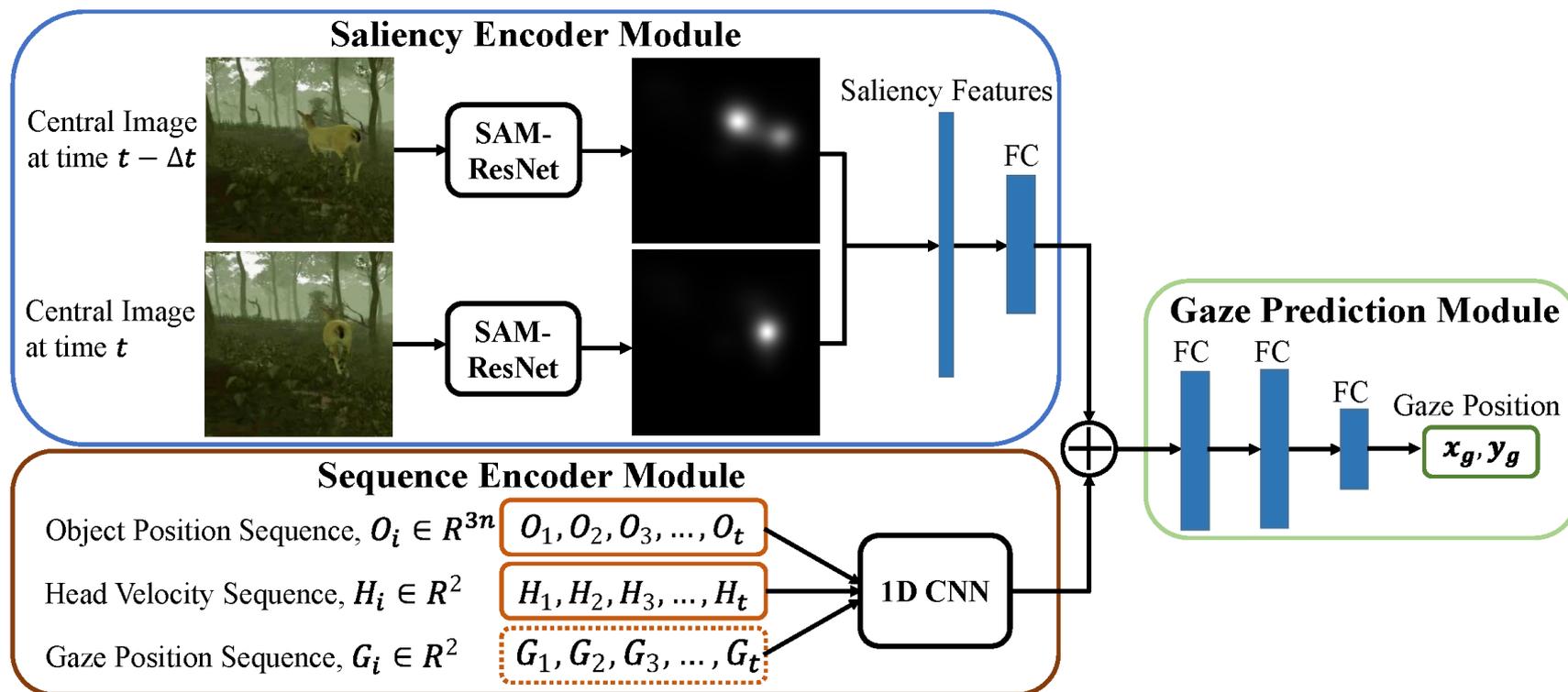
注视与动态物体相关性



注视位置与动态物体位置的相关性
左：水平方向；右：垂直方向

注视与动态物体具有相关性

DGaze模型



DGaze模型框架

模型评估

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

DGaze 模型与其他方法在动态和静态数据集上的预测效果

DGaze模型在动态和静态数据集上均有显著的提升

预测效果

Realtime gaze prediction results

任务驱动虚拟场景注视预测



[Hu et al. TVCG 2021 (Best Journal Nominees)]

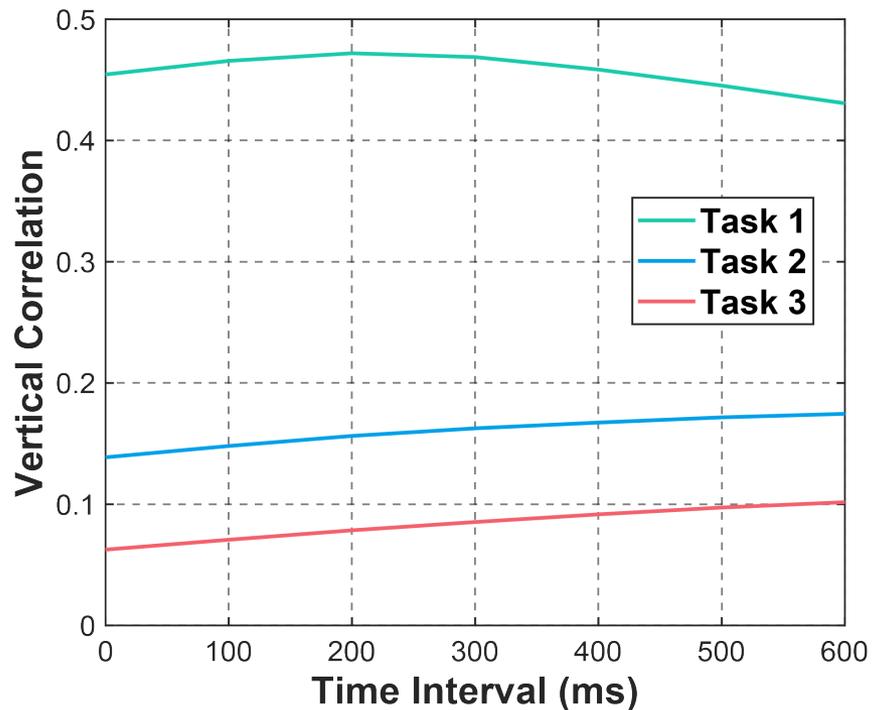
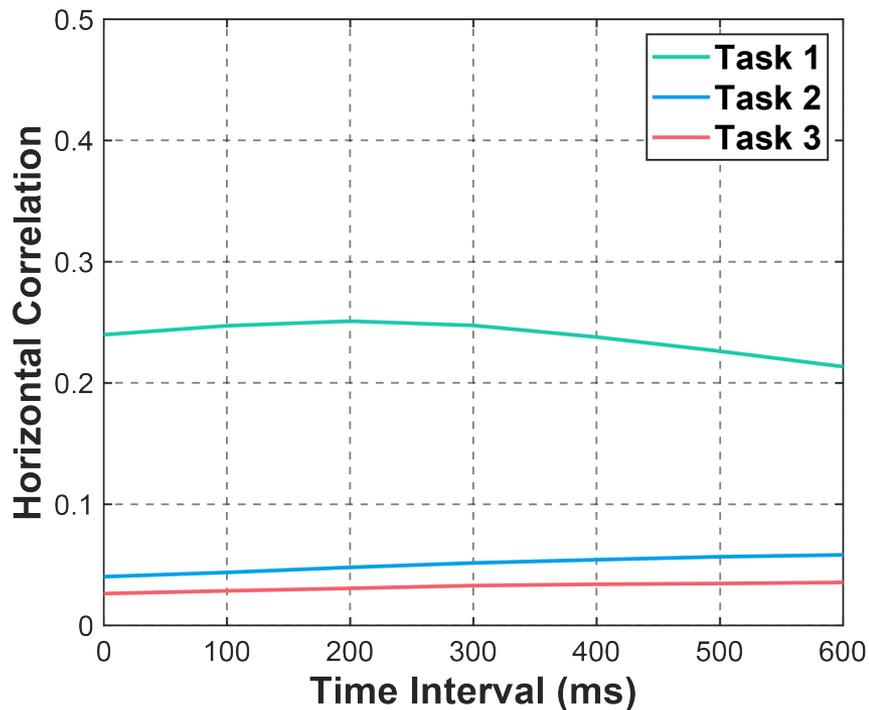
注视数据收集

- ▶ 参与者：27名用户（15男，12女，年龄17-32）
- ▶ 实验场景：四个任务驱动虚拟场景
- ▶ 实验设备：HTC Vive头盔、眼动仪
- ▶ 实验过程：视觉搜索任务
- ▶ 数据：场景内容、任务相关物体信息、眼睛运动、头部运动



实验场景

注视-任务相关性

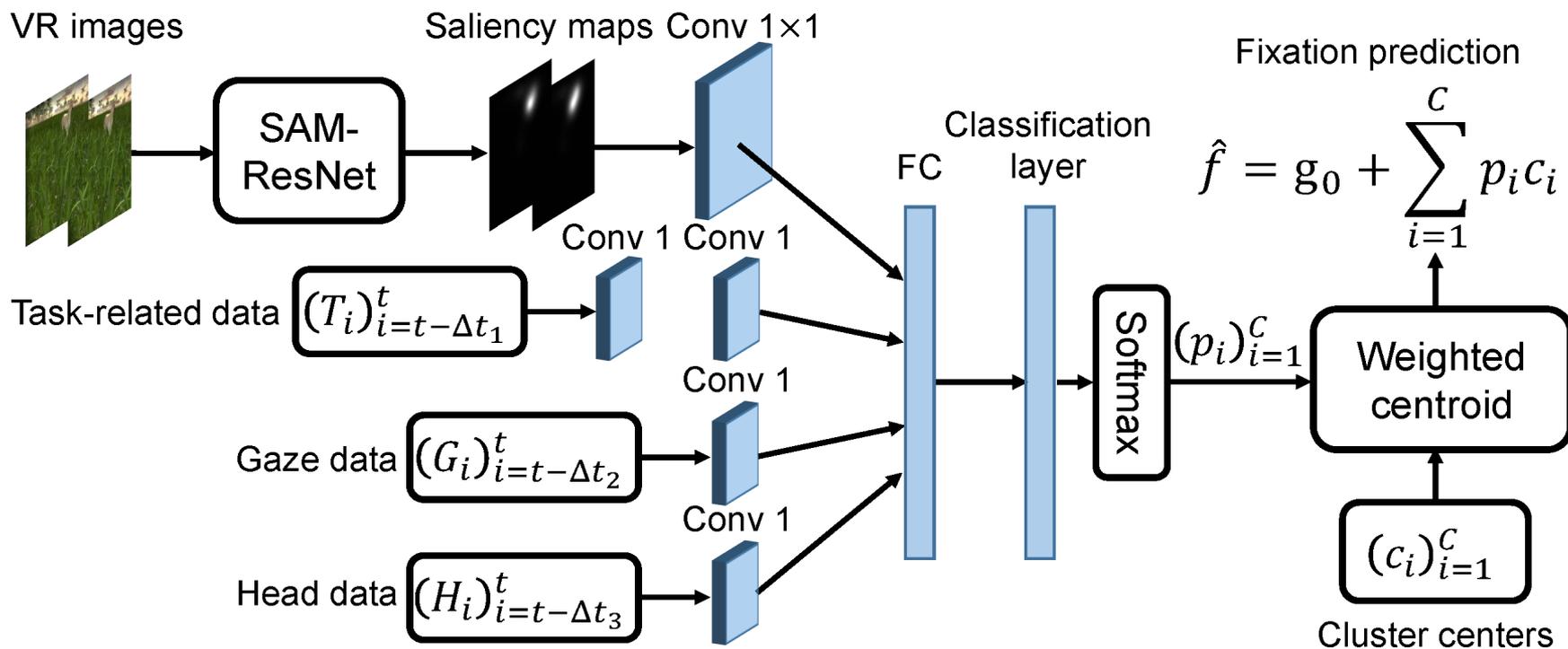


注视位置与任务相关物体的相关性

左：水平方向；右：垂直方向

注视与任务具有相关性

FixationNet模型



FixationNet模型框架

注视预测网络

$$\hat{f} = g_0 + \sum_{i=1}^c p_i c_i$$

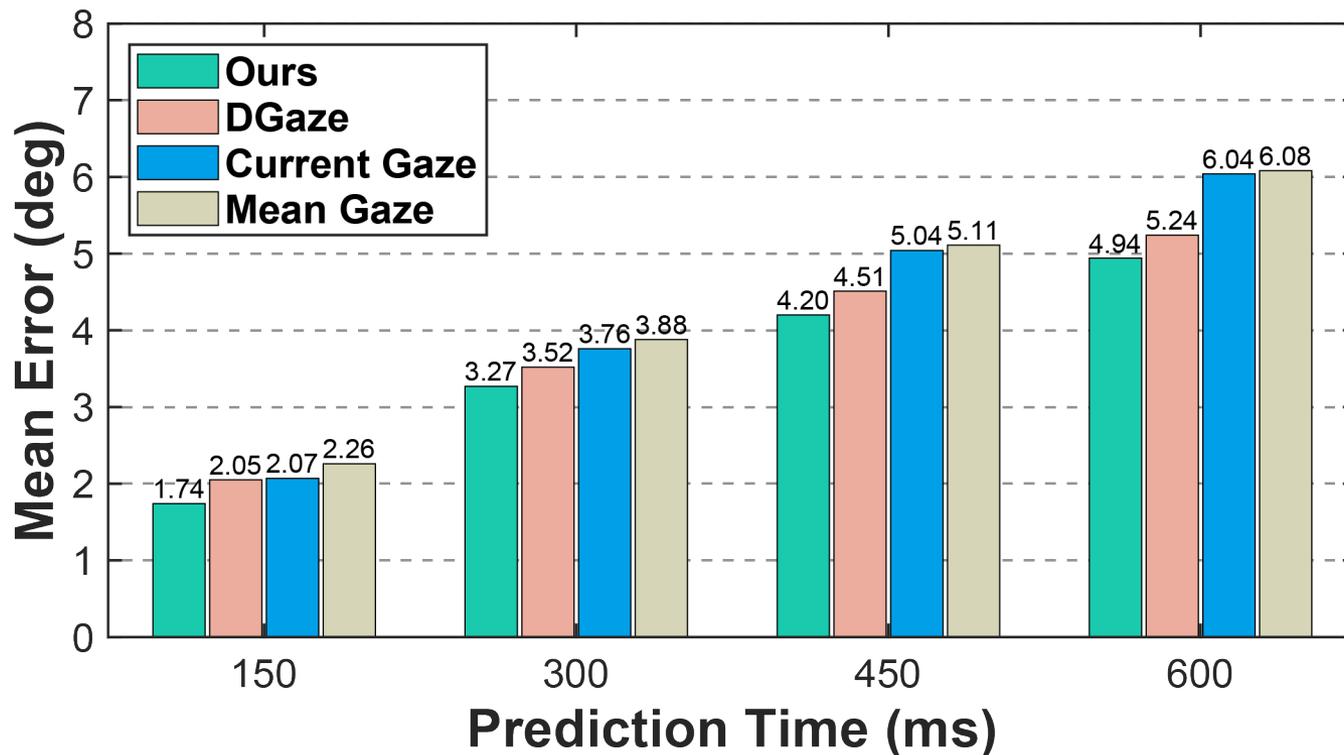
\hat{f} : 预测的未来时刻注视位置

g_0 : 用户在当前时刻的注视位置

c_i : 聚类中心的位置

p_i : 聚类中心的权重

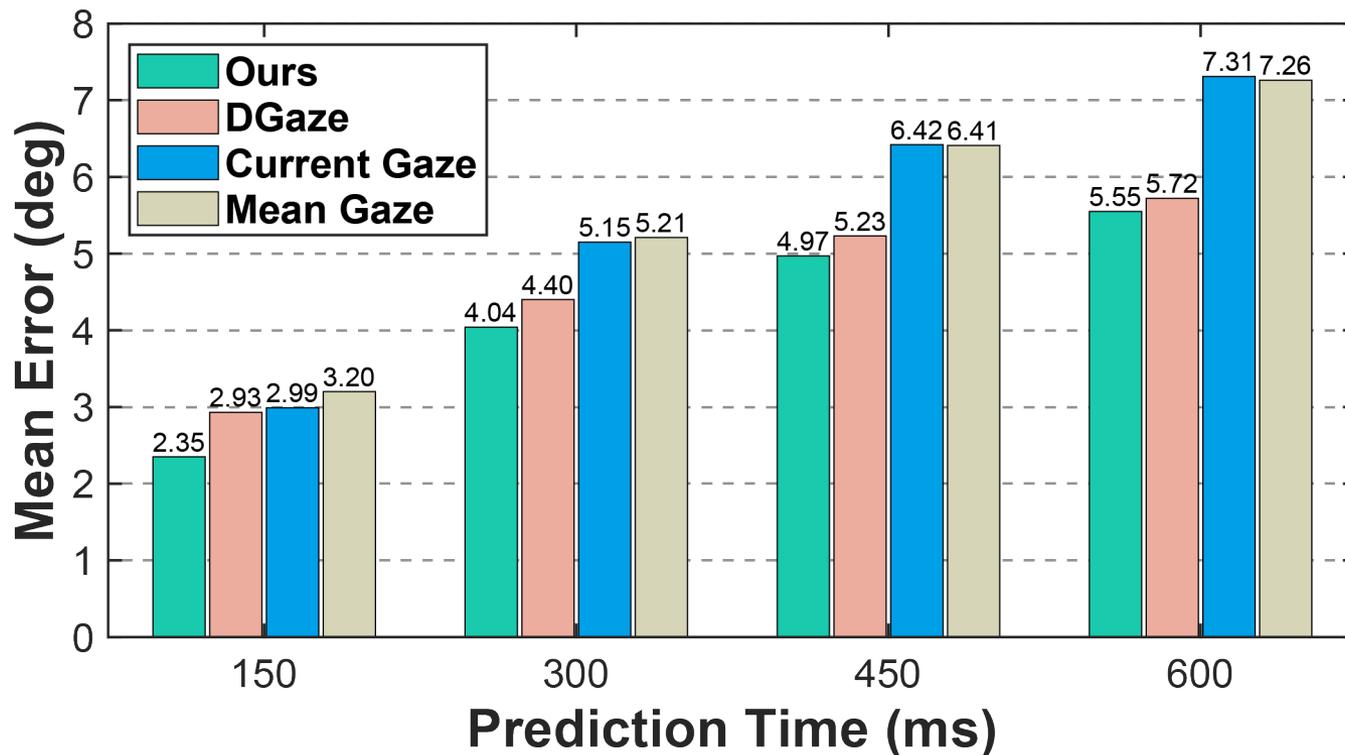
任务驱动场景评估



任务驱动场景中，不同时间间隔下模型的预测表现

FixationNet在不同预测时间下均具有最好的表现

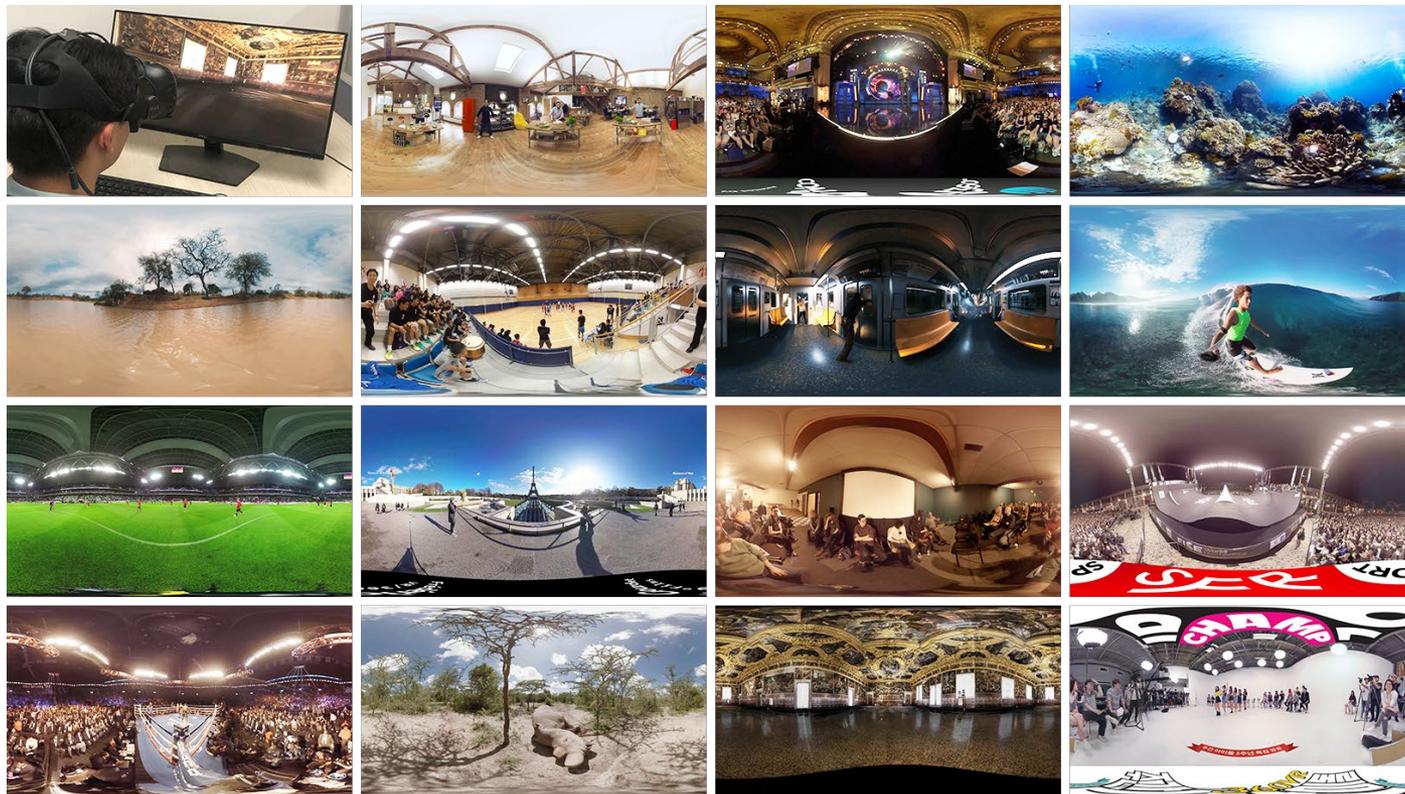
自由观察场景评估



自由观察场景中，不同时间间隔下模型的预测表现

在不同预测时间下，FixationNet都具有最佳的预测表现

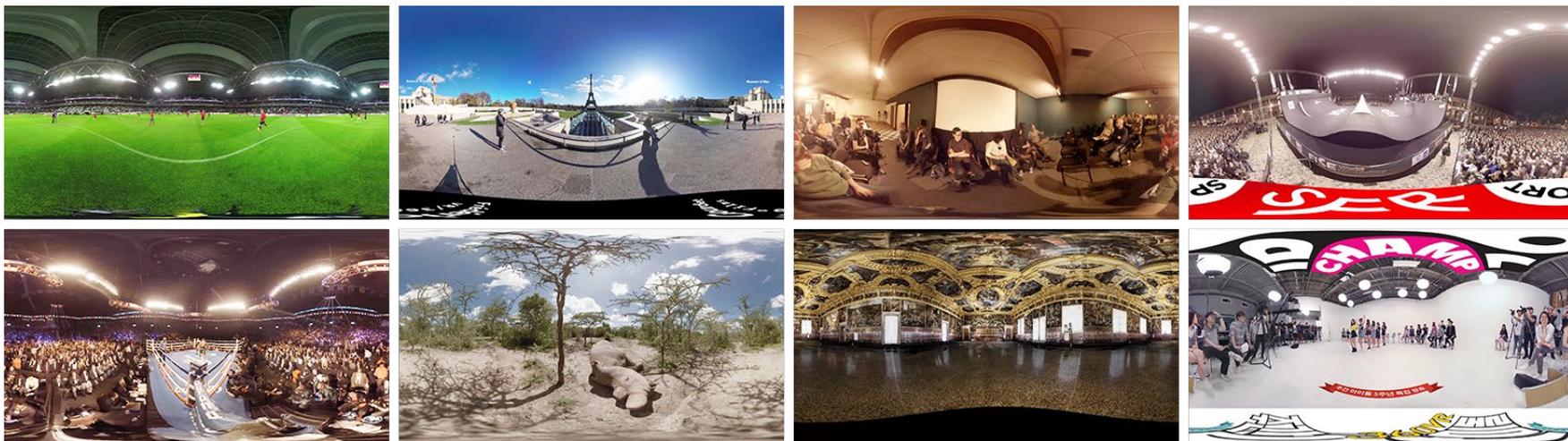
360度虚拟现实视频用户任务识别



[Hu et al. TVCG 2022]

数据收集

- ▶ 参与者：30名用户（18男，12女，年龄 $\mu = 24.5, \sigma = 5.0$ ）
- ▶ 实验场景：15个360度VR视频（时长150s）
- ▶ 实验设备：HTC Vive头盔、眼动仪
- ▶ 实验过程：自由观察、视觉搜索、显著性判断、物体追踪
- ▶ 数据：眼睛运动、头部运动、任务类别



实验场景

不同虚拟现实任务中用户眼睛运动的统计特征

		Free viewing	Visual search	Saliency	Track
Mean Fixation Duration	Mean	263.4 ms	339.5 ms	<u>241.2</u> ms	431.7 ms
	SD	25.6 ms	49.0 ms	24.3 ms	106.7 ms
Fixation Number Per Second	Mean	1.41	1.97	<u>1.22</u>	1.77
	SD	0.38	0.17	0.43	0.19
Mean Saccade Duration	Mean	633.2 ms	269.3 ms	<u>776.0</u> ms	241.1 ms
	SD	218.0 ms	69.2 ms	260.1 ms	56.2 ms
Saccade Number Per Second	Mean	1.03	1.20	<u>0.95</u>	1.01
	SD	0.17	0.18	0.19	0.24
Mean Saccade Amplitude	Mean	6.51°	4.73°	<u>8.56°</u>	5.40°
	SD	1.24°	1.05°	1.49°	1.58°
Fixation Distribution Dispersion	Mean	2.21E-6	2.25E-6	<u>7.08E-6</u>	2.50E-6
	SD	1.01E-6	1.18E-6	3.50E-6	1.57E-6

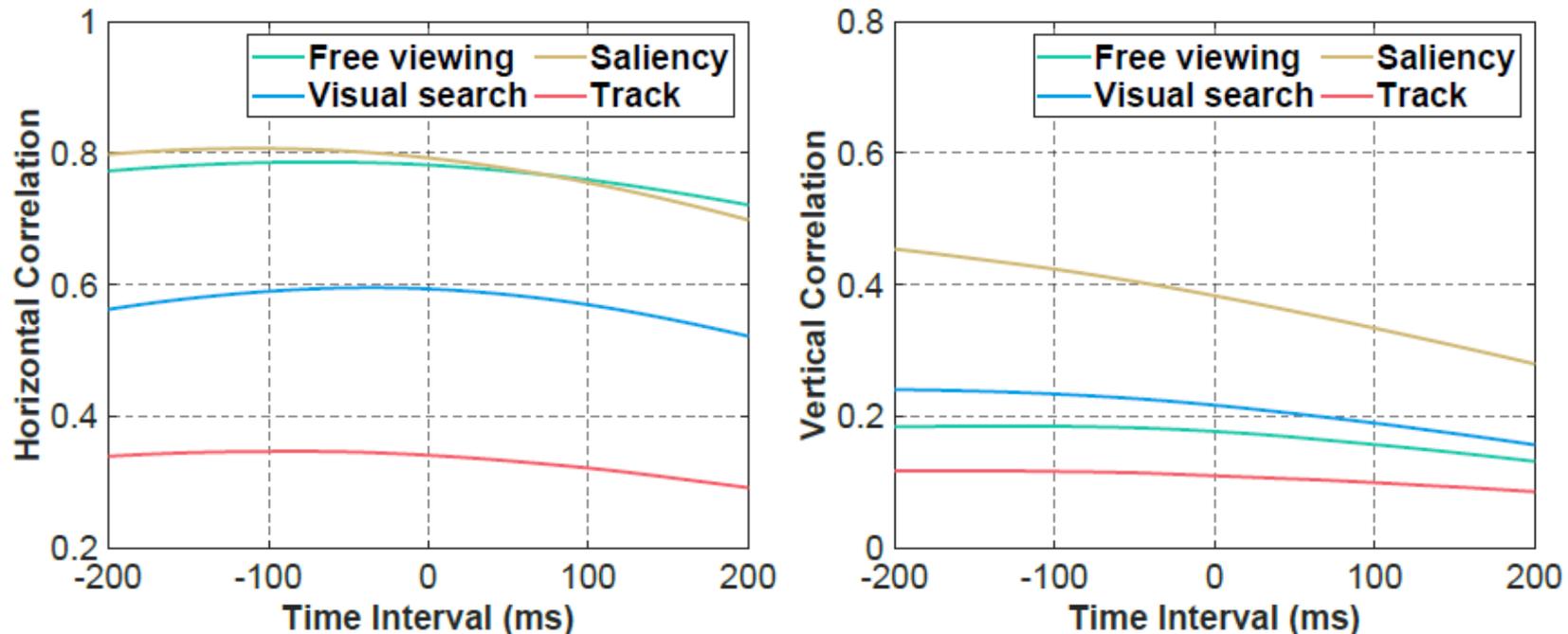
对表中的每一项特征值，如果两个任务之间的字体不相同，则表示这两个任务在该特征上存在着显著性的差异

不同虚拟现实任务中用户头部运动的统计特征

		Free viewing	Visual search	Saliency	Track
Mean Absolute Horizontal Velocity	Mean	22.7°/s	9.1°/s	<u>26.8°/s</u>	6.4°/s
	SD	4.3°/s	2.3°/s	4.4°/s	2.4°/s
Mean Absolute Vertical Velocity	Mean	2.9°/s	2.7°/s	<u>7.5°/s</u>	1.9°/s
	SD	0.6°/s	0.5°/s	1.4°/s	0.4°/s
Mean Absolute Horizontal Acceleration	Mean	182.6°/s²	140.4°/s ²	<u>203.5°/s²</u>	129.8°/s ²
	SD	29.4°/s ²	14.1°/s ²	23.9°/s ²	19.4°/s ²
Mean Absolute Vertical Acceleration	Mean	125.0°/s²	114.2°/s ²	<u>145.4°/s²</u>	109.4°/s ²
	SD	15.0°/s ²	11.1°/s ²	12.0°/s ²	11.6°/s ²
Velocity Distribution Dispersion	Mean	2.64E+4	6.95E+3	<u>2.39E+5</u>	3.12E+3
	SD	2.13E+4	7.98E+3	1.27E+5	4.35E+3

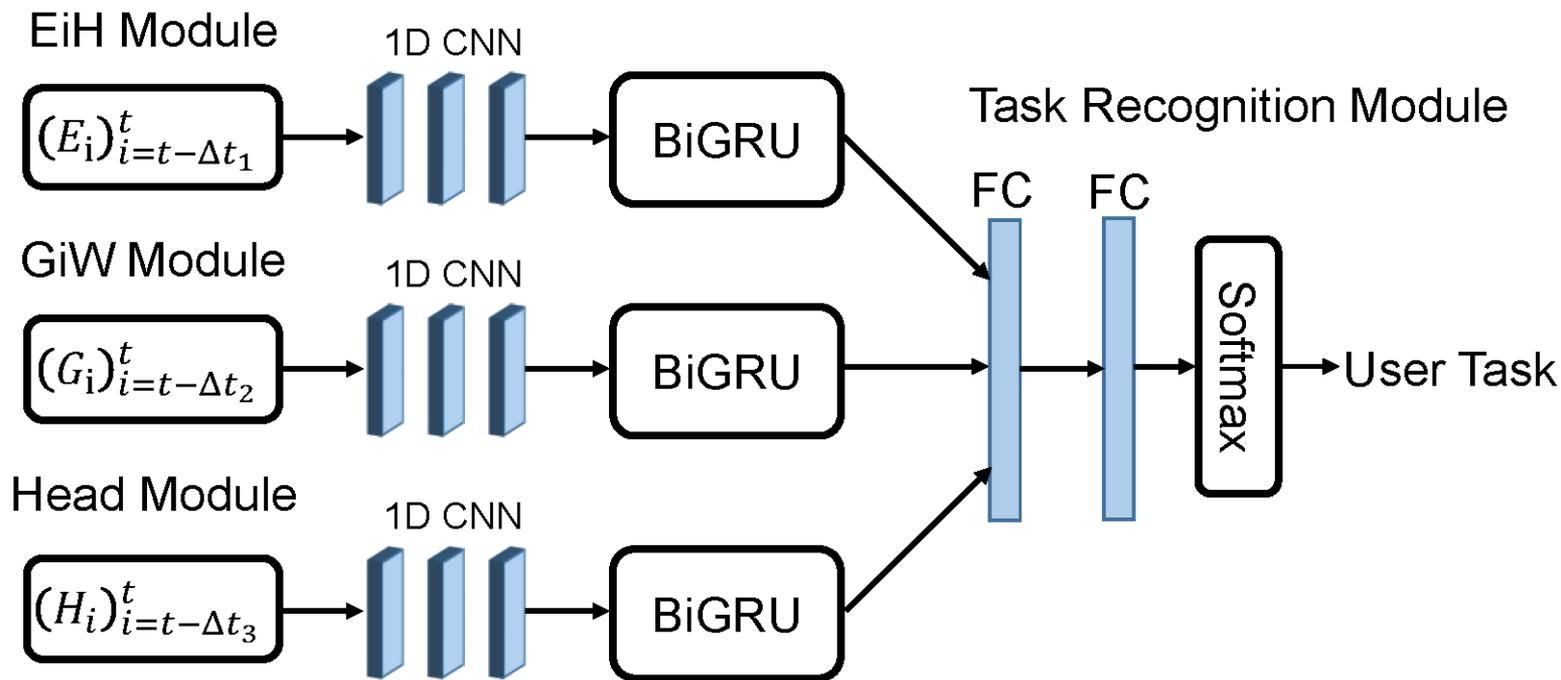
对表中的每一项特征值，如果两个任务之间的字体不相同，则表示这两个任务在该特征上存在着显著性的差异

不同虚拟现实任务中的眼动头动协调性



水平（左）和竖直（右）方向上的眼动头动协调性

EHTask模型



EHTask模型框架

虚拟现实场景模型评估

		Ours	LDA	SVM	BC	RFo	RFe
Cross-User	Window	84.4%	54.0%	54.3%	49.3%	<u>62.8%</u>	48.7%
	MV	97.8%	76.1%	75.3%	65.3%	<u>83.1%</u>	68.3%
Cross-Scene	Window	82.1%	53.8%	54.1%	49.0%	<u>62.6%</u>	48.3%
	MV	96.4%	74.2%	75.3%	64.4%	<u>83.6%</u>	72.2%

EHTask模型与其他方法在用户交叉和场景交叉下的预测表现

EHTask模型在用户交叉和场景交叉下的预测表现均有显著提升

现实场景模型评估

	Ours	LDA	SVM	BC	RFo	RFc
Window	61.9%	39.0%	37.9%	36.3%	<u>44.1%</u>	36.1%
MV	87.7%	60.0%	46.2%	40.0%	60.0%	<u>64.6%</u>

EHTask模型与其他方法在现实场景数据集中的表现

EHTask在现实场景中的表现显著好于其他方法

▶ 静态虚拟场景注视预测

基于眼动头动协调性的静态虚拟场景注视预测模型

▶ 动态虚拟场景注视预测

基于卷积神经网络的动态虚拟场景注视预测模型

▶ 任务驱动虚拟场景注视预测

基于学习的任务驱动虚拟场景注视预测模型

▶ 虚拟现实环境用户任务识别

基于学习的虚拟现实环境用户任务识别模型

- ▶ 探索其他因素对用户视觉注意的影响
声音、用户的手部运动、用户的身体运动、用户的心理状态
- ▶ 研究更加复杂的虚拟现实场景中的用户视觉注意
其他类型动态场景：交通工具、行人
其他类型任务场景：文本编辑、装配、阅读、记忆
- ▶ 研究虚拟现实环境中其他的视觉注意问题
心理负荷估计、心理意象重建、视觉意图预测、注意状态分类
- ▶ 探索其他环境中的用户视觉注意
增强现实环境、混合现实环境

发表的学术论文

- [1] Zehui Lin, Xiang Gu, Sheng Li, **Zhiming Hu**, Guoping Wang. Intentional Head-Motion Assisted Locomotion for Reducing Cybersickness, TVCG, 2021. (**CCF A**, Accepted)
- [2] **Zhiming Hu** et al. EHTask: Recognizing User's Task from Eye and Head Movements in Immersive Virtual Reality, TVCG, 2021. (**CCF A**)
- [3] **Zhiming Hu** et al. Research progress of user task prediction and algorithm analysis (in Chinese), Journal of Graphics, 2021.
- [4] **Zhiming Hu**. Eye Fixation Forecasting in Task-Oriented Virtual Reality, Proc. IEEE VR Abstracts and Workshops (VRW), 2021. (**CCF A**, Poster)
- [5] **Zhiming Hu** et al. FixationNet: Forecasting Eye Fixations in Task-Oriented Virtual Environments, TVCG, 2021. (**CCF A**, **TVCG Best Journal Nominees Award**)
- [6] **Zhiming Hu**. Gaze Analysis and Prediction in Virtual Reality, Proc. IEEE VR Abstracts and Workshops (VRW), 2020. (**CCF A**, Poster)
- [7] **Zhiming Hu** et al. DGaze: CNN-Based Gaze Prediction in Dynamic Scenes, TVCG, 2020. (**CCF A**)
- [8] **Zhiming Hu** et al. Temporal continuity of visual attention for future gaze prediction in immersive virtual reality, Virtual Reality & Intelligent Hardware, 2020.
- [9] **Zhiming Hu** et al. SGaze: A Data-Driven Eye-Head Coordination Model for Realtime Gaze Prediction, TVCG, 2019. (**CCF A**)

荣誉奖励

- 2021年12月，**国家奖学金**
- 2021年12月，北京大学三好学生
- 2021年3月，**TVCG最佳期刊论文提名奖**
- 2020年12月，北京大学二等奖学金
- 2020年12月，北京大学三好学生
- 2020年8月，国家留学基金委奖学金
- 2020年5月，**北京大学校长奖学金**
- 2019年12月，廖凯原奖学金
- 2019年12月，闳材奖学金
- 2019年12月，北京大学三好学生
- 2017年12月，闳材奖学金

提问环节

谢谢