

# 数字人姿态协调性研究

胡志明

北京大学

[jimmyhu@pku.edu.cn](mailto:jimmyhu@pku.edu.cn)

<https://zhiminghu.net/>

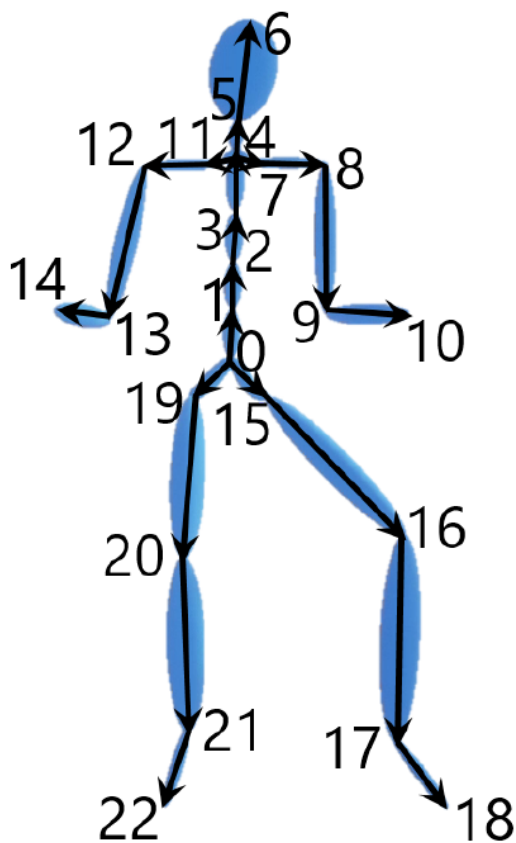
学历：北京大学计算机学院2022届博士

工作：德国斯图加特大学博士后

荣誉奖励：

- 2022年6月，北京大学优秀毕业生
- 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（国内首次）
- 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，北京大学三好学生

- ▶ 数字人概念
- ▶ 基于眼动头动协调性的注视预测模型
- ▶ 动态虚拟现实环境中的注视预测模型
- ▶ 任务驱动虚拟现实环境中的注视预测模型
- ▶ 虚拟现实环境中的用户任务识别模型



- |    |             |    |             |
|----|-------------|----|-------------|
| 0  | Root        | 11 | L. Collar   |
| 1  | Chest       | 12 | L. Shoulder |
| 2  | Chest2      | 13 | L. Elbow    |
| 3  | Chest3      | 14 | L. Wrist    |
| 4  | Chest4      | 15 | R. Hip      |
| 5  | Neck        | 16 | R. Knee     |
| 6  | Head        | 17 | R. Ankle    |
| 7  | R. Collar   | 18 | R. Toe      |
| 8  | R. Shoulder | 19 | L. Hip      |
| 9  | R. Elbow    | 20 | L. Knee     |
| 10 | R. Wrist    | 21 | L. Ankle    |
|    |             | 22 | L. Toe      |

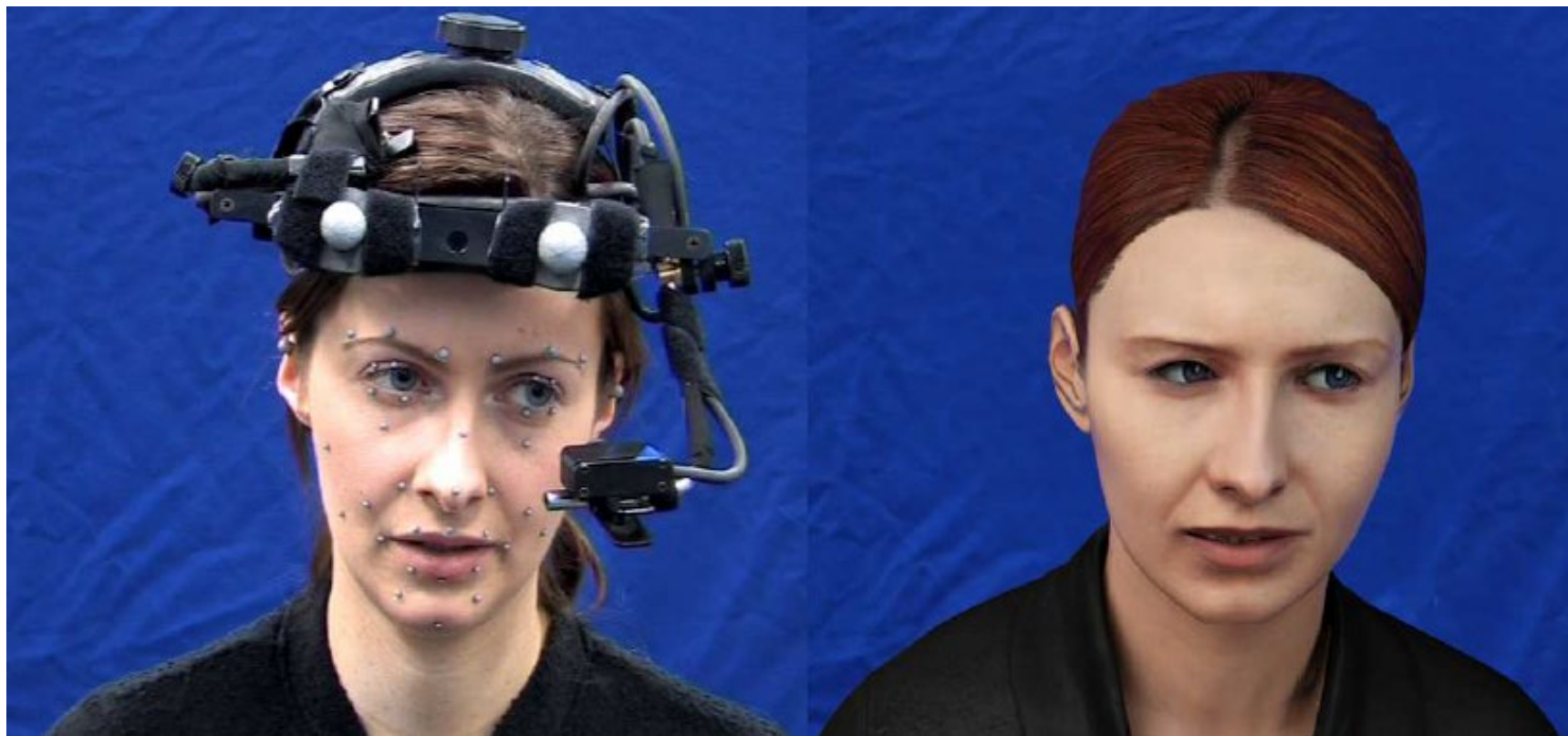
[Bhattacharya et al. 2021]



[Randhavane et al. 2019]



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



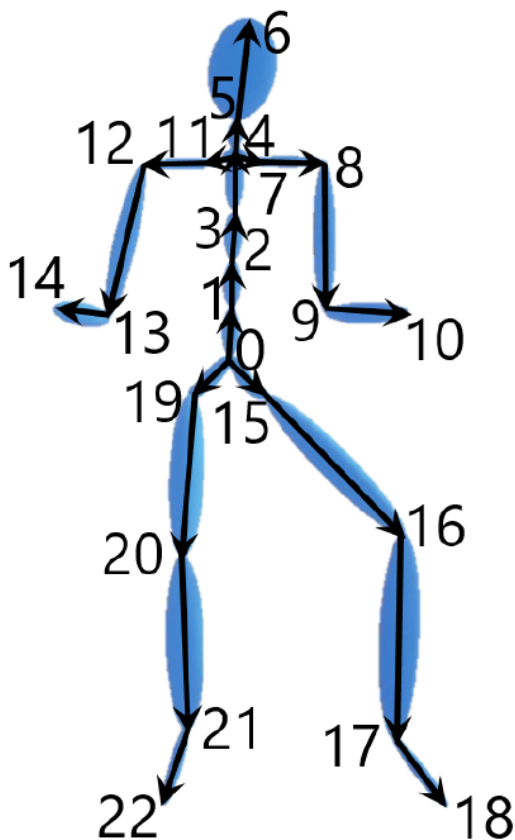
[McDonnell et al. 2015]

## 数字人相关研究问题

- ▶ 数字人形象生成（图片、交互输入）
- ▶ 数字人驱动（文本、语音、交互输入）
- ▶ 数字人真实感绘制（光照、纹理、阴影）



## 数字人姿态协调性



- |    |             |    |             |
|----|-------------|----|-------------|
| 0  | Root        | 11 | L. Collar   |
| 1  | Chest       | 12 | L. Shoulder |
| 2  | Chest2      | 13 | L. Elbow    |
| 3  | Chest3      | 14 | L. Wrist    |
| 4  | Chest4      | 15 | R. Hip      |
| 5  | Neck        | 16 | R. Knee     |
| 6  | Head        | 17 | R. Ankle    |
| 7  | R. Collar   | 18 | R. Toe      |
| 8  | R. Shoulder | 19 | L. Hip      |
| 9  | R. Elbow    | 20 | L. Knee     |
| 10 | R. Wrist    | 21 | L. Ankle    |
|    |             | 22 | L. Toe      |

[Bhattacharya et al. 2021]

## 数字人姿态协调性

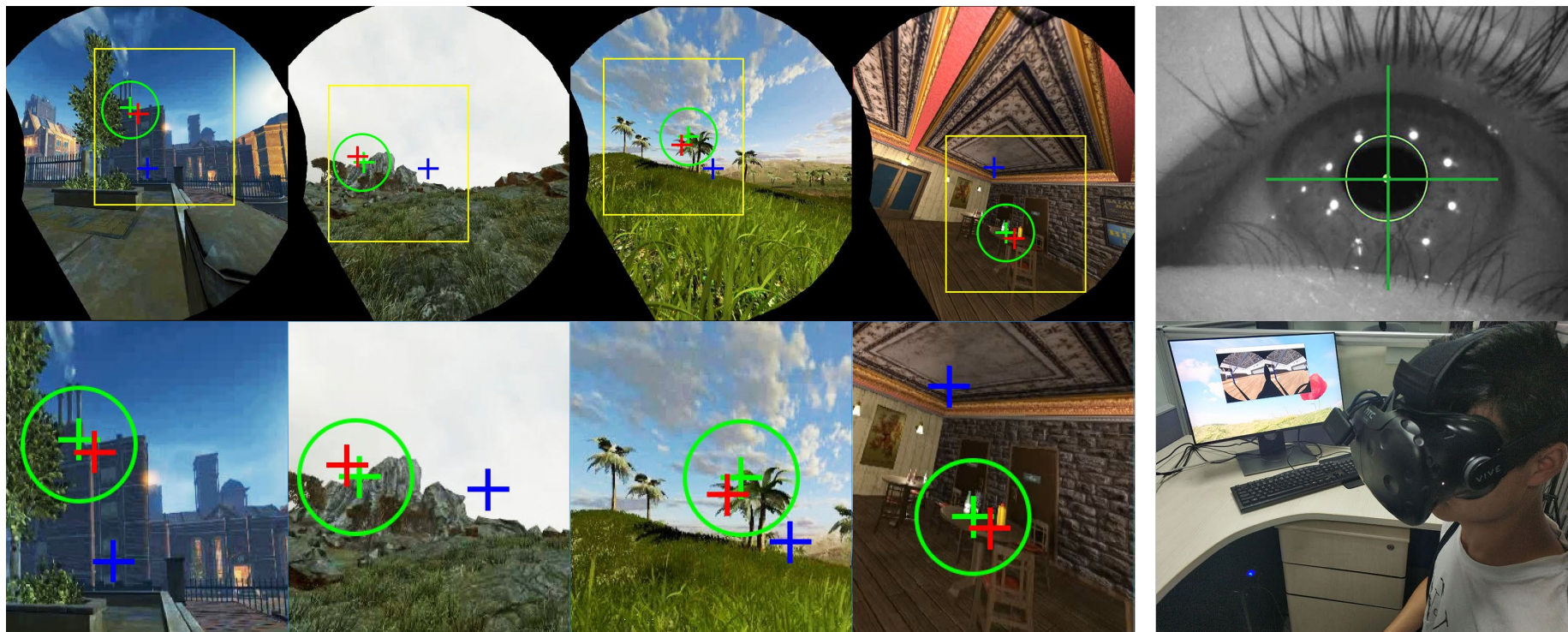
▶ 眼动-头动协调性

▶ 眼动-身体运动协调性

▶ 头动-手动协调性

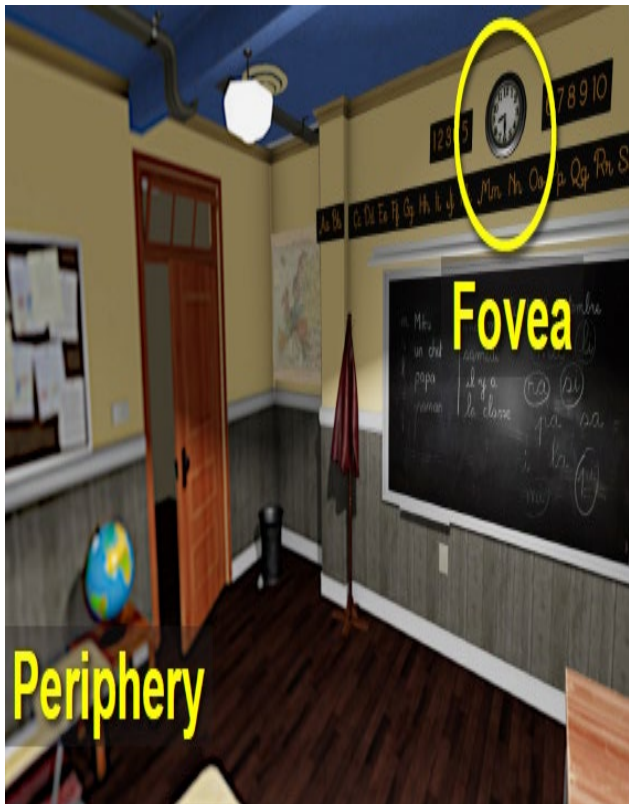
▶ 身体-步态协调性

## 静态虚拟场景注视预测

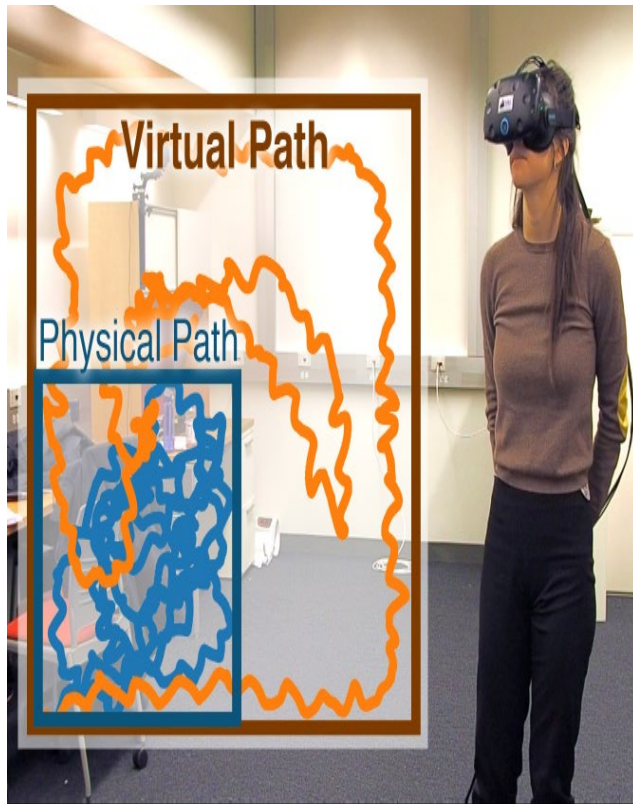


[Hu et al. TVCG 2019]

## 注视预测应用



注视点渲染  
[Patney et al. 2016]

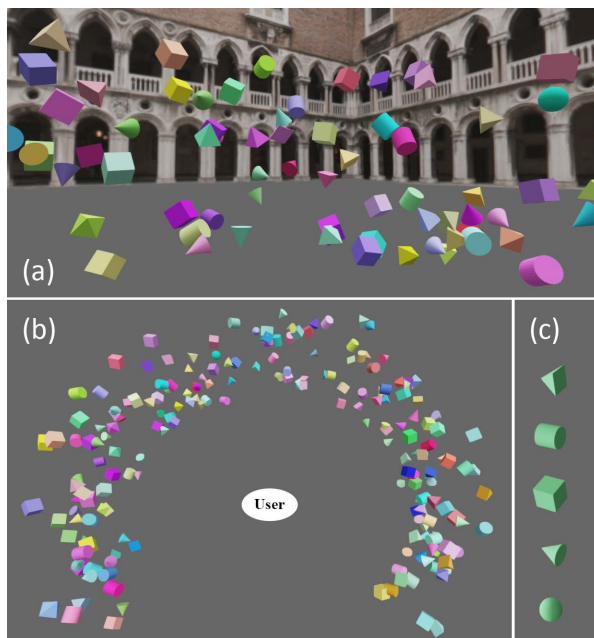


重定向行走  
[Sun et al. 2018]



场景布局优化  
[Alghofaili et al. 2019]

## 注视预测应用



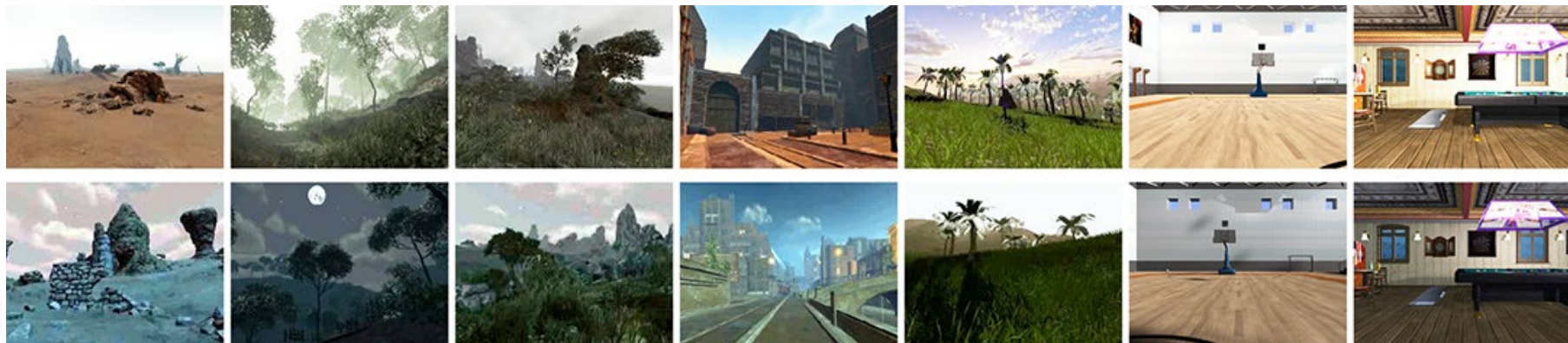
VR内容设计  
[Sitzmann et al. 2018]

注视引导  
[Grogorick et al. 2017]

细节层次管理  
[Lee et al. 2009]

## 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、注视位置、头部运动



实验场景

## 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

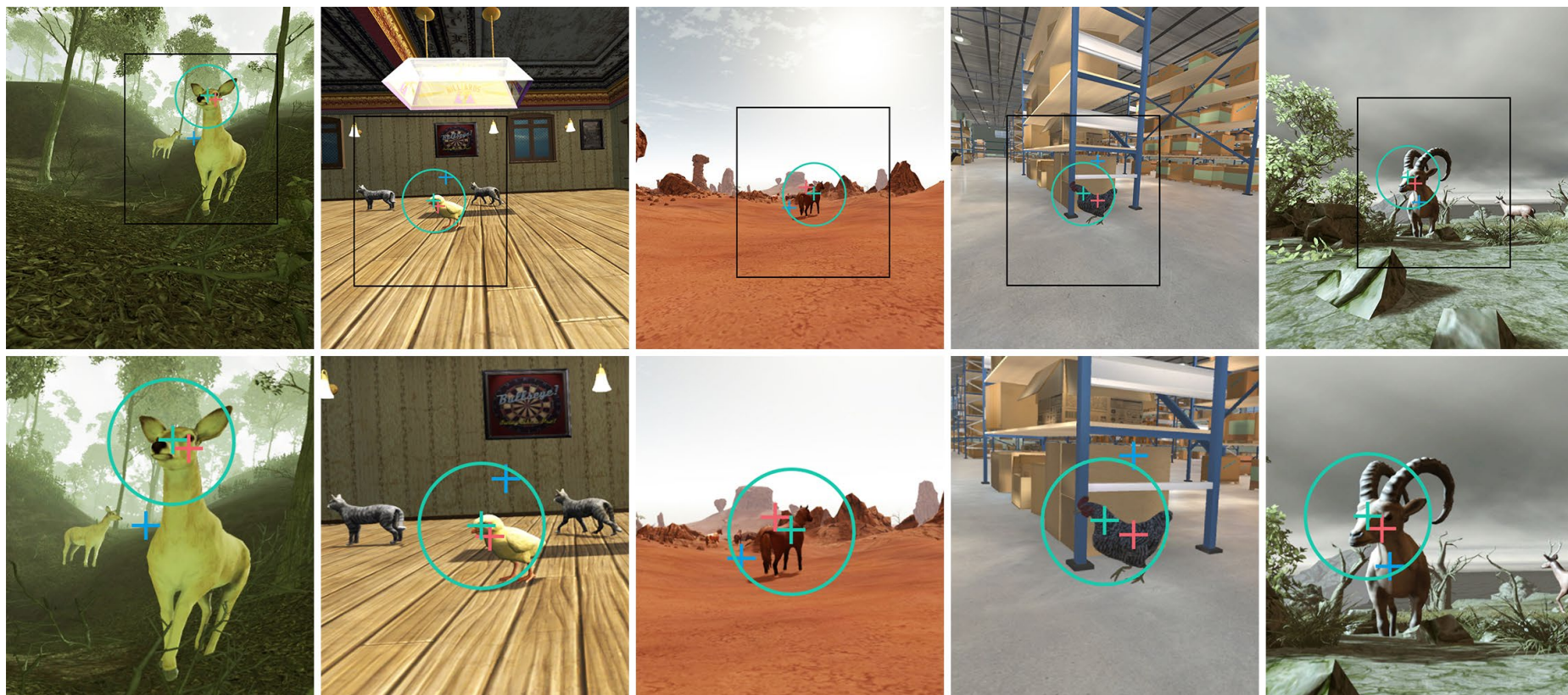
	SGaze	Mean	Center	Saliency
Mean Error	<b>8.52°</b>	10.93°	11.16°	21.23°

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



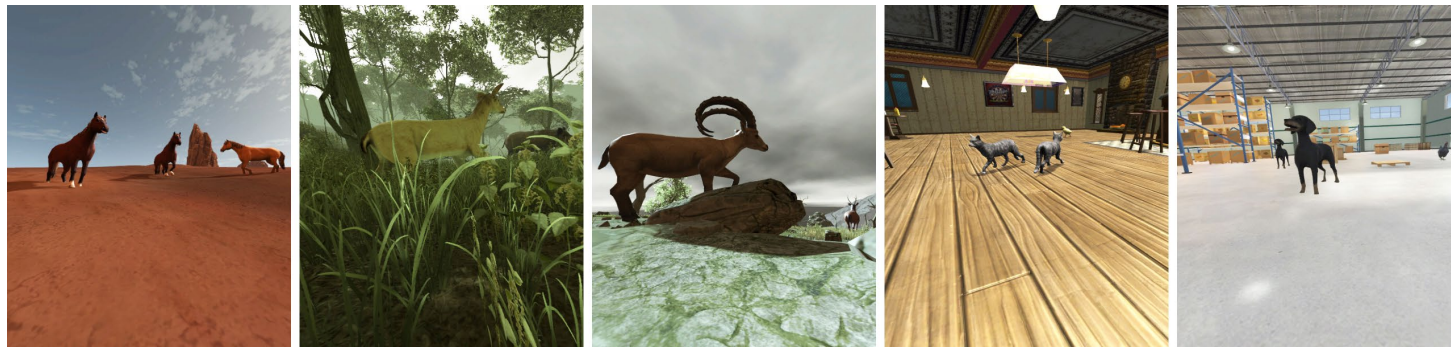
## 动态虚拟场景注视预测



[Hu et al. TVCG 2020]

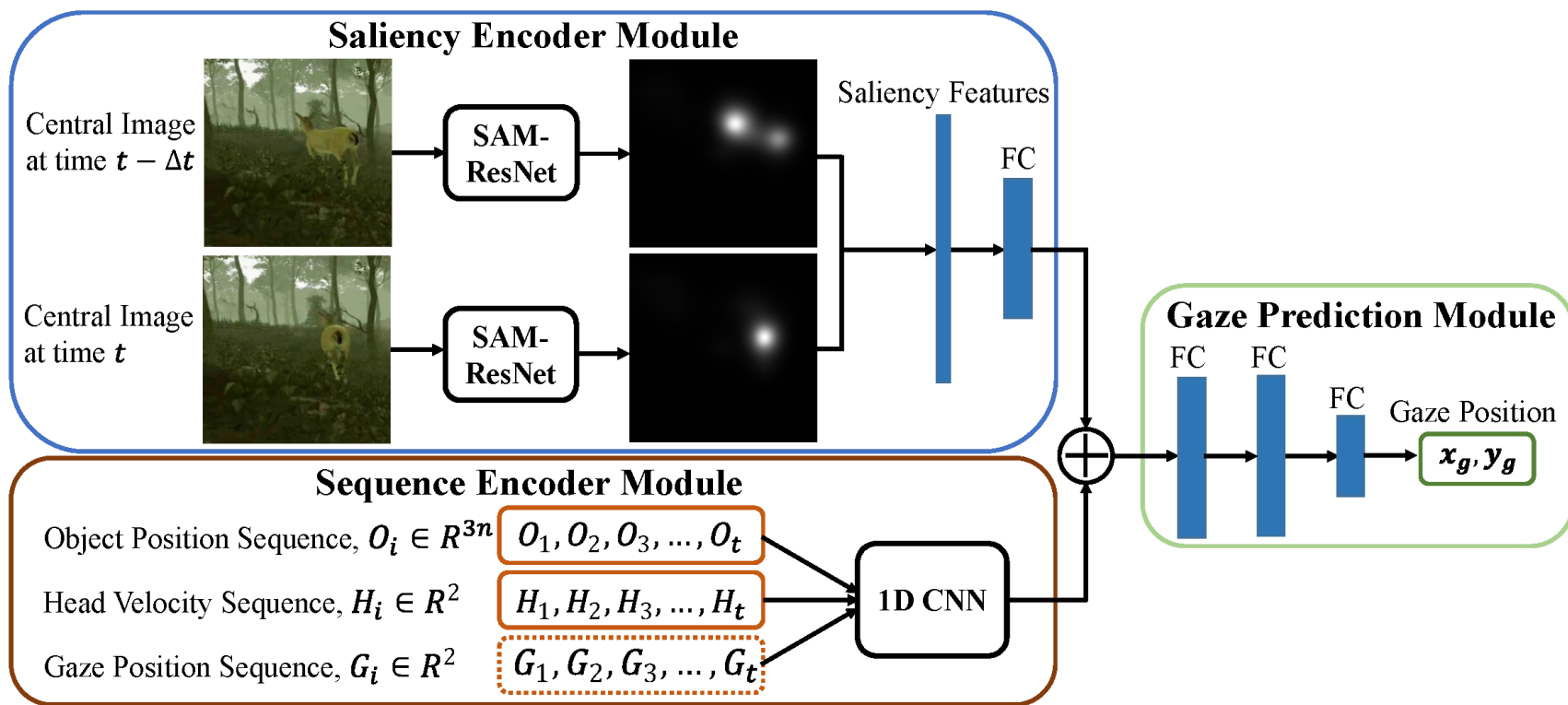
## 数据收集

- ▶ 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、注视位置、头部运动



实验场景

## DGaze模型



DGaze模型框架

## 模型评估

		DGaze	SGaze	Mean	Center	Object
Dynamic	Mean Error	<b>7.11°</b>	9.11°	10.04°	12.46°	13.25°
Static	Mean Error	<b>7.71°</b>	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*

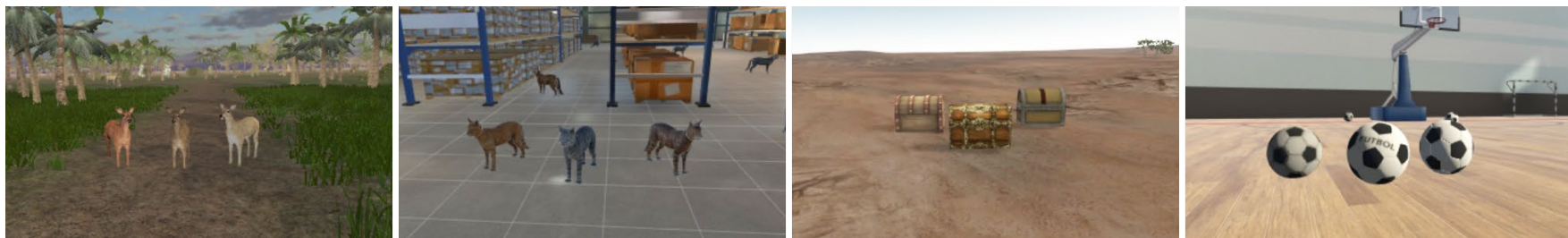
## 任务驱动虚拟场景注视预测



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

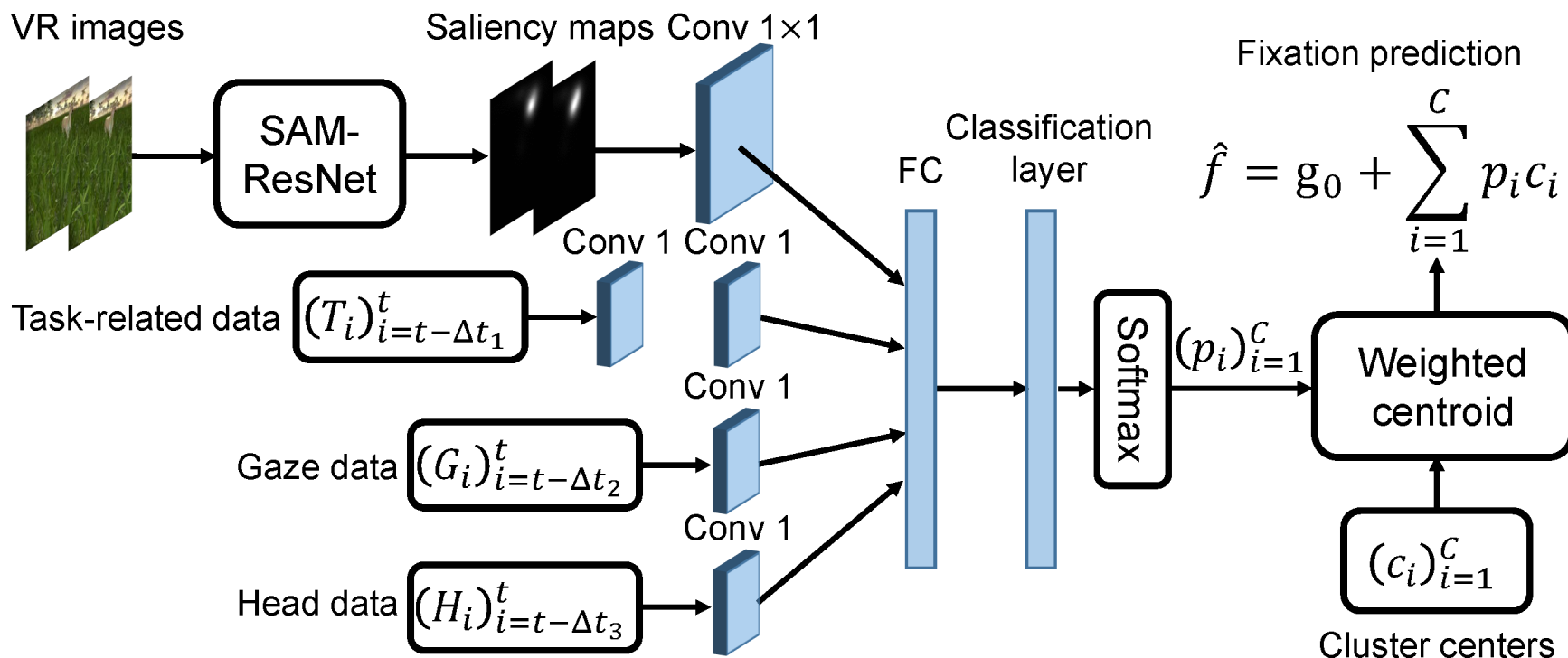
## 数据收集

- ▶ 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、注视位置、头部运动



实验场景

## FixationNet模型



FixationNet模型框架



## 注视预测网络

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

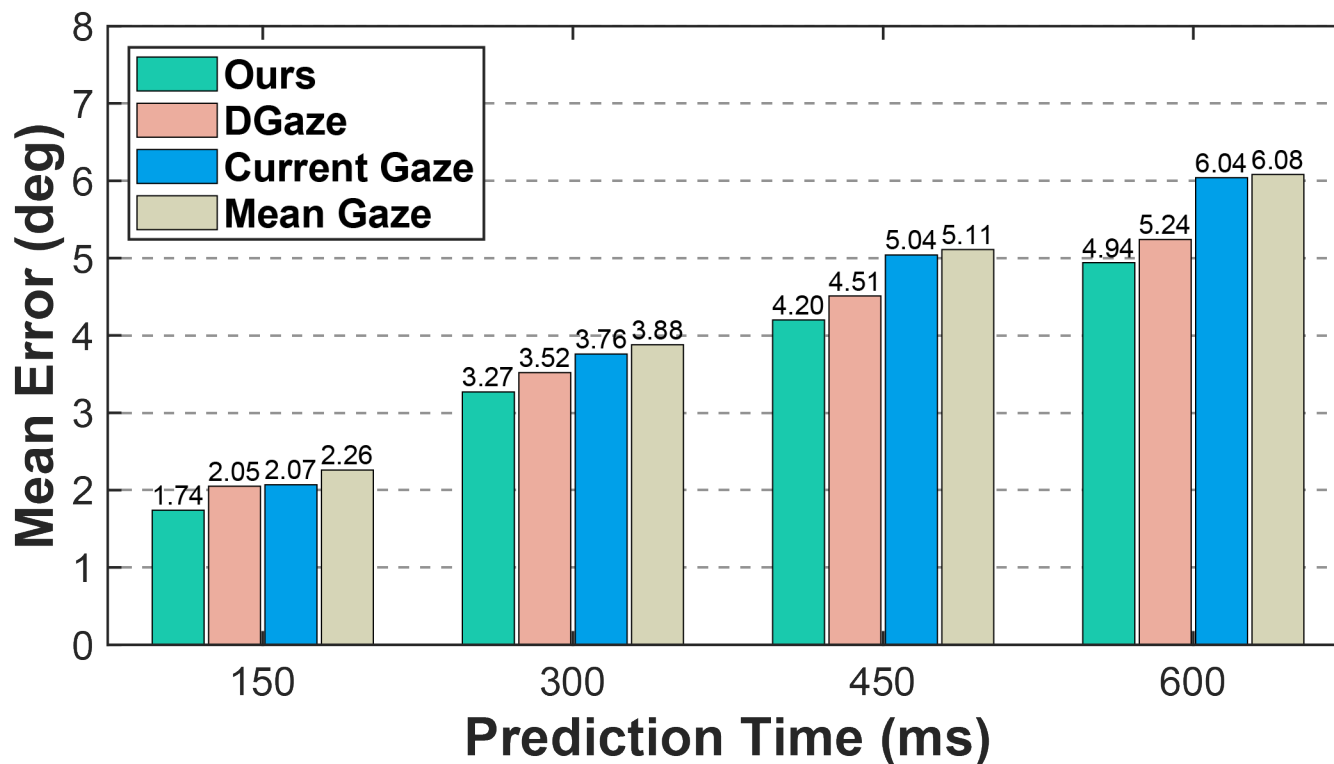
$\hat{f}$ : 预测的未来时刻注视位置

$g_0$ : 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

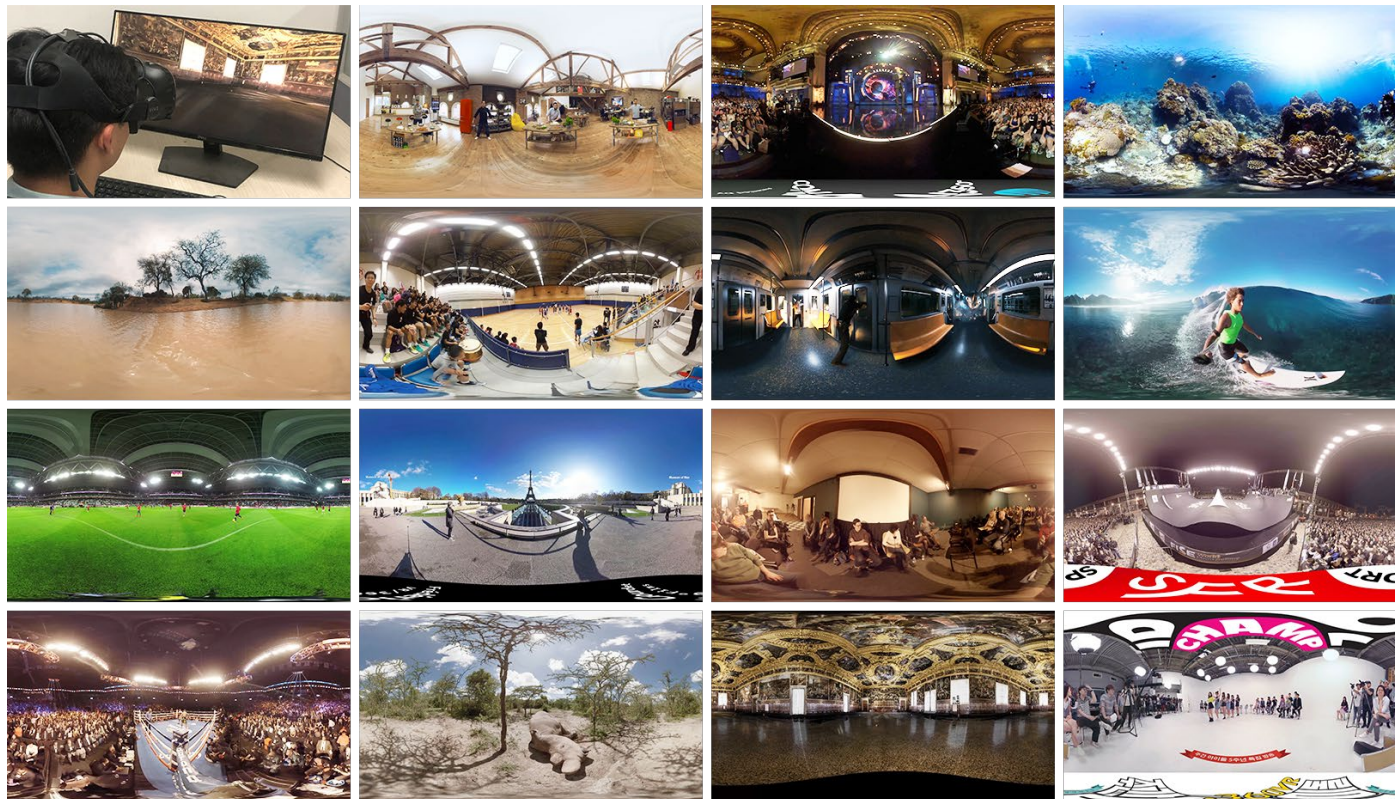
## 模型评估



任务驱动场景中，不同时间间隔下模型的预测表现

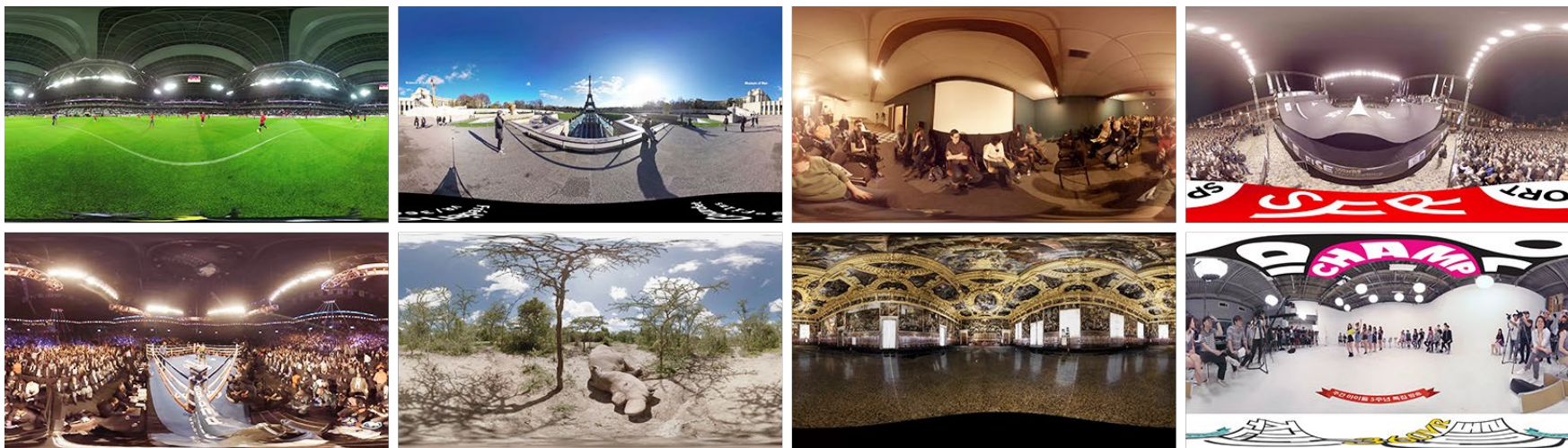
FixationNet在不同预测时间下均具有最好的表现

## 360度VR视频用户任务识别



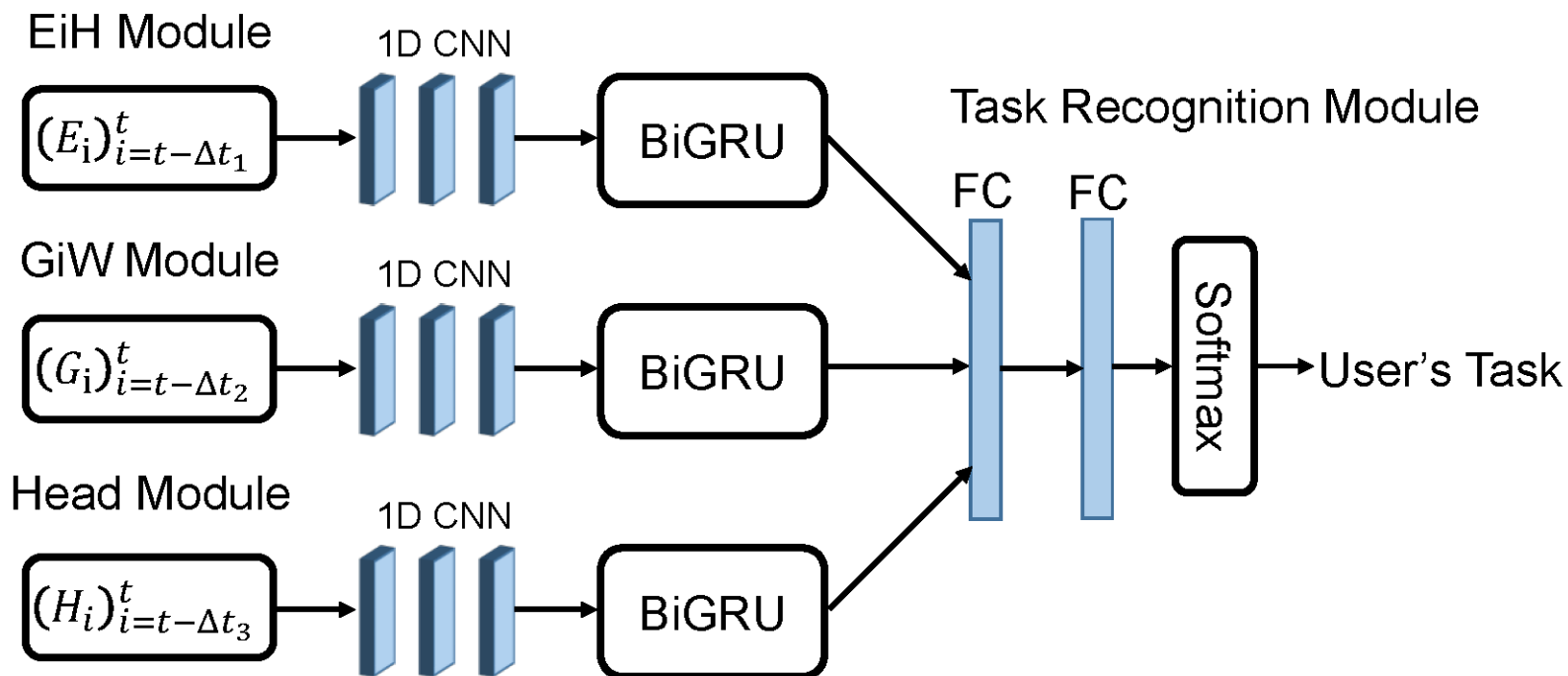
## 数据收集

- ▶ 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：任务类别、注视位置、头部运动



实验场景

## EHTask模型



EHTask模型框架

## 模型评估

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

## 公众号：小侦探的书桌

- ▶ 北大硕士的德国读博之路
- ▶ 北大女孩的外企生活
- ▶ 互联网工作两年后的思考



谢谢