

# 轻量化空间视觉注意力量化研究初探

## ——以高校科研建筑空间为例

肖婉凝<sup>1</sup>, 邓巧明<sup>2</sup>

**摘要:** 在建筑设计领域, 视觉注意力的量化分析对于评估设计方案的有效性至关重要。尽管已有研究通过眼动追踪技术探讨了建筑空间对用户视觉注意力的影响, 但高昂的设备成本和操作复杂度限制了这一技术的广泛应用。本研究开发一种基于 JavaScript 的开源库 WebGazer 的轻量化眼动追踪网页平台, 支持通过电脑摄像头进行空间视觉图像的眼动追踪实验, 并自动导出注视点坐标数据。进一步, 结合 Python 的 pandas、matplotlib、sklearn 等库进行注视点热力图、视觉焦点聚类、视线方向等可视化分析。研究为建筑设计等领域提供一种低成本、易于访问的视觉注意力量化工具, 并发挥 python 快速可视化数据分析的优势, 构建相对完整且便捷的空间视觉注意力量化的 workflow。

**关键词:** 科研建筑空间; 空间视觉分析; 注意力量化; 眼动追踪技术; 轻量化网页平台开发

**Keywords:** Research Architectural Space; Spatial Visual Analysis; Attention Quantification;

Eye-tracking Technology; Lightweight Web Platform Development

**资助项目情况:** 国家自然科学基金项目 No:51978268; 广东省自然科学基金项目

No:2023A1515010758

## 引言

建筑设计领域中, 对空间视觉注意力的量化研究具有重要意义。量化空间视觉注意力, 可以在设计过程中识别最吸引人的设计元素以及兴趣点区域、设计减少压力的环境、提升用户的空间感知和认知体验、实现数据驱动设计<sup>[1]</sup>。近年来, 许多学者利用眼动追踪技术开展空间视觉注意力的分析研究, 为建筑设计带来更准确、客观的分析方法。此外, 眼动追踪技术也应用于建筑工程和施工行业中, 包括设计评估与优化、用户体验分析、安全性评估、教育和培训、建筑遗产保护、施工过程监控等方面<sup>[2]</sup>。但高昂的设备成本、严谨的实验环境使其使用门槛较高, 在非正式研究与设计环境中, 探索能高效、轻量地进行粗略视觉注意力分析的研究方法, 对眼动追踪技术的应用及推广有重要价值。

## 1 研究背景

### 1.1 空间视觉注意力量化研究方法

<sup>1</sup> 肖婉凝, 华南理工大学建筑学院, 亚热带建筑与城市科学全国重点实验室, 硕士研究生

<sup>2</sup> 邓巧明(✉), 华南理工大学建筑学院, 亚热带建筑与城市科学全国重点实验室, 教授, dengqm@scut.edu.cn

基于已有研究，空间视觉注意力的量化大致可总结为三个步骤。首先，确定视觉指标来记录参与者的眼动数据，评价视觉注意力。根据研究目标的差异，其指标包含注视指标、扫视指标和瞳孔指标<sup>[3]</sup>、视线停留时间和观察路径<sup>[4]</sup>、视觉注意力热点、注视点和注视持续时间<sup>[5]</sup>等。此外，通过改变注意窗口形状和扩散范围，还能测量参与者在不同视觉任务中的注意力分布<sup>[6]</sup>。其次，在视觉图像分析上，可使用真实建筑场景或虚拟建模场景的图像数据，将显著性模型与实际人类观察者的眼动数据进行比较<sup>[7]</sup>。对于复杂的环境图像，可先进行语义分割来比较各区域差异。最后，采用图像可视化方式呈现量化数据。通过结合脑电信息和眼动数据，生成建筑空间的热图，直观地展示用户偏好，帮助设计师优化建筑空间<sup>[8]</sup>。

### 1.2 不同眼动追踪技术的设备要求与使用现状

根据实验设备和成本差异，常用的眼动追踪设备有以下几种。桌面眼动追踪器通常安装在显示器前或显示器上，主要用于实验室环境中的静态视觉任务研究，成本相对较低。头戴式眼动追踪器是用户佩戴在头上的设备，适用于动态环境中的视觉任务研究，成本较高但提供更丰富的数据。视频型眼动追踪器精度高，可以在多种环境中使用，使用红外光源照亮眼睛，通过摄像头记录眼球运动，并利用算法计算视线位置，其成本中等，适合需要高精度数据的研究。电生理眼动追踪器设备通过电极测量眼睛周围的电场变化，其设备适合便携使用<sup>[9]</sup>。目前，虽然眼动追踪技术在视觉注意力量化方面有较好成效，但因具有较高的使用门槛，使用较为局限。

### 1.3 基于 WebGazer 的轻量化眼动追踪技术应用相关研究

2016年，网络眼动追踪库 WebGazer.js 由布朗大学(Brown University)的帕普萨基(Alexandra Papoutsaki)领导的研究团队首次公布。WebGazer.js 利用 JavaScript 和 web 技术，通过分析来自计算机摄像头的视频数据，实现实时眼动追踪，为广泛和多样化的参与者群体进行眼动追踪研究提供了可能。

相较于其他常见的眼动追踪设备，网页眼动追踪技术的核心优势是开源免费、应用灵活，允许开发者对其进行修改以适应特定的研究或应用需求。其次易于集成，可轻松集成到网页，不需要专业眼动追踪硬件即可通过普通的网络摄像头收集眼动数据，目前已经在心理学、市场研究、用户界面设计等多个研究领域中广泛应用。该技术与其他常见眼动追踪设备的对比分析见表1。

表1 集成 WebGazer.js 的眼动追踪技术与其他眼动追踪设备的对比分析表（表格来源：作者自绘）

类型	精度	硬件要求	使用场景	舒适度	成本
集成WebGazer的眼动追踪技术 <sup>[10]</sup>	中等 约 4-5 度	无专用硬件，基于网络摄像头	网络浏览器环境，轻量化实验	高	低
桌面眼动追踪器 <sup>[11]</sup>	高 约 0.5 度	安装在显示器上的专业硬件，包含摄像头和、红外光源	静态环境，适用于用户体验测试等研究	中	高
头戴式眼动追踪器 <sup>[11]</sup>	高 约 1 度	内置摄像头和传感器的头戴设备	动态环境，适用于 VR 应用和实地研究	中	高
视频型眼动追踪器 <sup>[11]</sup>	高 约 1 度	固定安装在工作台上的高帧率摄像头和计算设备	实验室环境，适用于精确的视觉研究	高	高
电生理眼动追踪器 <sup>[11]</sup>	最高 约 0.001 度	电极或电极帽等生理测量设备	实验室环境，适用于高度精确的实验研究	低	最高

WebGazer.js 的主要局限是精度不足问题，包括空间和时间两个层面。空间上，WebGazer.js 在估计注视位置时的空间偏差相较于传统的红外眼动追踪设备较大，估计的目标注视点与实际之

间的平均空间偏差大约为 4-5 度<sup>[10]</sup>。受环境和设备的影响，时间上，存在眼动追踪效果可能会出现延迟<sup>[12]</sup>等局限。可以通过实验过程中调整电脑摄像头角度、维持光环境稳定、设置校准实验等方法改善这些不足。

## 2 研究方法与框架

### 2.1 研究方法

实验基于一台 8GB RAM 且自带摄像头的笔记本电脑，通过搭建在线网页播放一系列空间视觉图像，内置眼动追踪收集用户的眼动数据，分析用户在观看图像时的注视点位置和注视时间。

网页采用 html 结构，头部部分包含网页的元数据和外部资源链接、网页的标题、定义网页的视觉样式等，主体部分包含展示图片的区域、间歇期间提示的区域和加载与执行脚本的区域。图片展示通过标签引入，控制其样式和展示方式。间歇页面在每张图片展示后显示“请答题”的提示页面，便于插入自定义实验问卷。脚本部分加载了 WebGazer.js 库，并且包含了核心的 JavaScript 代码，用于控制网页的交互逻辑，包括图片的切换、校准点的显示与用户交互、眼动数据的收集与导出等。

为削弱 WebGazer.js 的局限对实验结果的影响，研究进行眼动精确性的改良。首先，在实验前设置眼动校准。研究表明 WebGazer.js 使用注视估计算法来预测用户的注视位置在计算机屏幕上的具体位置，因此可利用鼠标点击行为来训练其模型，以改进对用户注视位置的估计<sup>[13]</sup>。后进行误差判断，将注视位置坐标和校准点坐标的平均偏差距离和屏幕像素值进行对比。经初步测试，在光线相对稳定的室内环境中，设定平均误差阈值在 8%-10%时，校准成功率可接近 100%。在满足实验要求的前提下，宽松的误差阈值能降低环境的要求、提升眼动校准的效率。其次，结合优化用户校准流程，提供清晰的指导和反馈能提升眼动识别的准确性。此外，实验全程可通过摄像框的实时监测用户头部位置，保证 WebGazer.js 的有效工作。再者，优化实验界面和交互设计，并设置合理的实验时长和浏览节奏，避免注意力分散和视觉疲劳对实验结果产生影响。最后，可以在用户界面上提供实时的注视点反馈，帮助用户调整其注视行为，减少实验误差。

### 2.2 研究框架

实验 workflow 可分为前期准备、眼动校准、正式实验、数据分析四大部分，研究框架如图 1 所示。前期准备时需要确认电脑摄像头完好，且环境光线均匀，场景图片大小一致并按顺序与 html 文件放置在同一文件夹中，用浏览器打开确认其能正常开启摄像头。进入眼动校准后受试者需保证头部在摄像框内，根据页面提示完成校准，正式实验时尽量不要改变头部位置。正式实验中为增加实验灵活性插入间歇页面，可自定义插入问卷答题或访谈等环节。在观看完所有图像后会导出 csv 格式数据，利用 python 进行数据的快速可视化分析，探究空间兴趣点的大致区域。

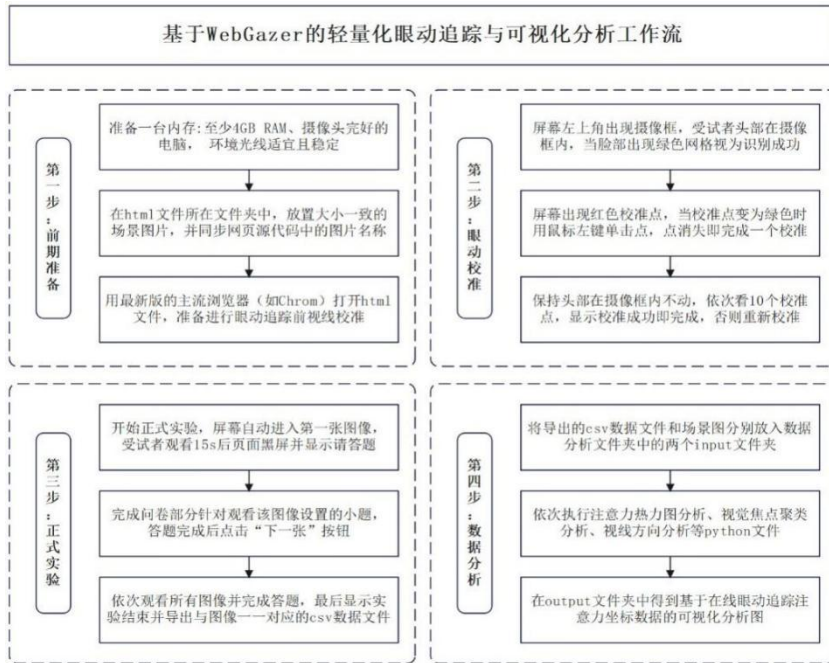


图1 研究框架（图片来源：作者自绘）

### 3 基于空间场景图像分析的眼动实验及量化分析

#### 3.1 实验对象和目的

实验对象是选址于华南理工大学五山校区南区，由7栋科研楼组成的科研建筑空间的设计方案。实验选取由AI控制性生成的10张空间场景图像，空间场景、功能和位置如图2所示。进行眼动实验的目的是量化分析科研建筑中空间要素和信息要素的兴趣点，对AI生成式设计的结果进行空间视觉方面的量化评估。



图2 空间场景、功能和位置（图片来源：作者自绘）

#### 3.2 眼动实验

眼动实验包含前期准备、眼动校准、正式实验三部分。

### 3.2.1 前期准备

实验地点选取在光线均匀的图书馆室内空间，光源包含顶部天窗自然光和室内灯光。实验设备为 NVIDIA GeForce RTX 4060 (8GB RAM) 且摄像头完好的笔记本电脑。网页 html 文件与 10 张像素与分辨率相同的空间场景 png 图片均存放于 C 盘的一个文件夹中。打开网页后左上角呈现摄像框，人脸在框内显示绿色即表示合格，显示红色表示不合格（见图 3）。研究招募 20 名受试者依次参与实验，为控制实验变量，20 名受试者均佩戴近视眼镜。

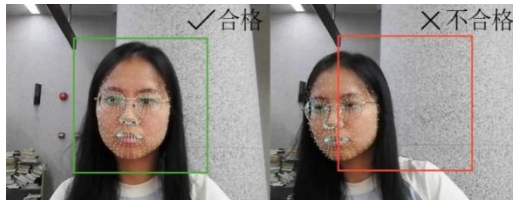


图 3 摄像框内的人脸位置合格性判定（图片来源：作者自绘）

### 3.2.2 眼动校准

为确保眼动追踪结果的准确性，正式实验开始前首先进行眼动校准测试。测试共设置 10 个校准点，前 5 个为固定点位，在屏幕的四角和中心点位置，后 5 个为随机点位。各个校准点在屏幕范围内依次出现，受试者需要注视当前出现的校准红点，当 2 秒后校准点由红色变成绿色时即可用鼠标左键单击，随后绿色点消失，下一个红色校准点显现。每次点击校准点时系统将自动记录用户实际的注视位置坐标和校准点坐标之间的误差，最后计算所有点的平均误差。由于该实验空间图像中各兴趣点分布较为分散，故误差阈值可以相对宽松：若平均误差阈值在屏幕对角线像素值的 10% 以内，则视为校准成功，可进入正式实验；若平均误差阈值大于 10% 则校准失败，重新开始校准，直至校准成功。测试结果显示，20 名受试者中，90% 的受试者一次即可通过校准测试，其余 10% 的受试者，在第二次也通过了校准测试。

### 3.2.3 正式实验

为增加网页代码的适配度和灵活性，图像展示逻辑是采用相对路径，将 png 图片和网页代码文件放在同一个文件夹中，完成并通过眼动校准后，即可按图片命名顺序进行依次展示，图像播放页面示例见图 4。播放图片张数取决于文件夹的图片数，每张图片播放时间为 15 秒，播放完后呈现答题间隔页面见图 5，用户在此期间可按需设置关于图像的问答题，实验要求受试者根据画面的空间和各个要素进行量表评价。答题无时长限制，完成后单击“下一张”按钮即可继续观看图像。全部图像播放完后，显示实验结束，并自动下载 csv 格式文件，每个文件对应每张图像，数据以图像左下角为原点，每 1 秒进行 2 次数据采集，分别记录受试者在 15s 内观察每张图像的注视时长和注视点坐标。

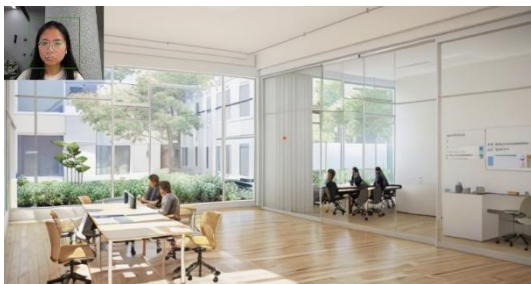


图 4 图像播放页面示例（图片来源：作者自绘）



图 5 答题间隔页面图（图片来源：作者自绘）

来源：作者自绘）

### 3.3 量化分析

#### 3.3.1 问卷量表分析

将 20 名受试者对于每张空间场景图像的空间和信息要素吸引力程度的量表评价进行数据整合，并利用 SPSS 26 进行描述性分析，分析结果见表 2。

表 2 问卷量表描述性统计分析（表格来源：作者自绘）

变量分类		量表统计值					
		均值	标准差	方差	偏度	峰度	
信息要素	平面	电视	4.15	0.08	0.01	0.32	-1.40
	信息媒介	标识	4.00	0.46	0.21	0.00	2.98
		海报	3.83	0.29	0.09	0.21	-0.55
		书	3.70	0.47	0.22	-0.95	-1.24
	立体	公告栏	3.60	0.60	0.36	0.39	-0.57
		模型	4.35	0.59	0.35	-0.21	-0.55
		设备	4.20	0.41	0.17	1.62	0.70
	互动场景	人与人交流场景	4.94	0.10	0.01	-1.28	0.54
		人机互动场景	4.93	0.16	0.03	-2.08	3.18
界面透明度		4.26	0.21	0.04	0.81	1.06	
空间要素	空间界面	界面材料色彩	3.40	0.60	0.36	1.25	0.78
		自然景观	3.22	0.32	0.10	0.79	1.25

由表可知，信息要素中，对受试者吸引力由高到低排序分别是：互动场景、立体媒介、平面媒介。其中互动场景均值最高，标准差和方差都较低，数据的离散程度很小，大多数评分都集中在均值附近，说明互动场景对人的视觉吸引力最大，在空间设计中容易激发人们的兴趣。具体来看，人与人交流场景（-1.28）和 人机互动场景（-2.08）都呈现出负偏度，表明这两个场景的评分分布向左偏，评分大多高于均值，表明大多数受试者对这两个场景非常感兴趣。此外，人与人交流场景（0.54）的峰度略高于 0，但接近正态分布，说明评分集中在均值附近。人机互动场景（3.18）的峰度较高，说明数据分布集中在一个较窄的范围内，并且评分集中在高值附近，显示出受试者高度一致地认为人机互动场景的视觉吸引力强。

其次，空间要素中，对受试者吸引力由高到低排序分别是：界面透明度、界面材料色彩和自然景观。界面透明度在三个变量中得分最高，且评价一致性较高，说明大多数受试者认为界面透明度会明显影响人的空间视觉感受。界面材料色彩虽然表现出正偏度（1.25），但其标准差较大，意味着受访者对材料色彩影响视觉吸引力的看法上分歧较大。自然景观的评价相对较低，但其偏度（0.79）和峰度（1.25）显示出分布在较高分数区域的集中度，故尽管整体得分不高，但大多数受访者的评价相对一致且偏向认可其对提升空间视觉吸引力具有一定积极影响。

#### 3.3.2 空间视觉图像可视化分析

将 20 名受试者对于每张空间场景图像的注视点坐标数据整合，利用 python 的 pandas、matplotlib、sklearn 等开源库对导出的坐标数据结合空间场景图像进行可视化分析。通过归一化方法定位注视点坐标在图像中的相对位置，可量化注视点的分布区域和数量。此外，结合采集坐标的时间顺序也可体现出受试者观看图像期间注视点位置的变化，综合分析各个空间的兴趣点区域。



### (1) 热力图分析

热力图 (Heatmap) 通过颜色的深浅或明暗来表示数据的数值大小或密度, 可展示多变量数据之间的关系或某个变量在不同区域或时间上的变化趋势。图 6 显示 10 张场景图像的热力图分析结果, 热力图由蓝到红体现受试者在各个场景图像中注视点由少到多的大致区域, 红色区域即吸引力最大的区域, 其主要分布在互动场景附近, 与量表分析结果基本一致。基于此, 可协助实验人员快速判断空间兴趣点的大致位置。

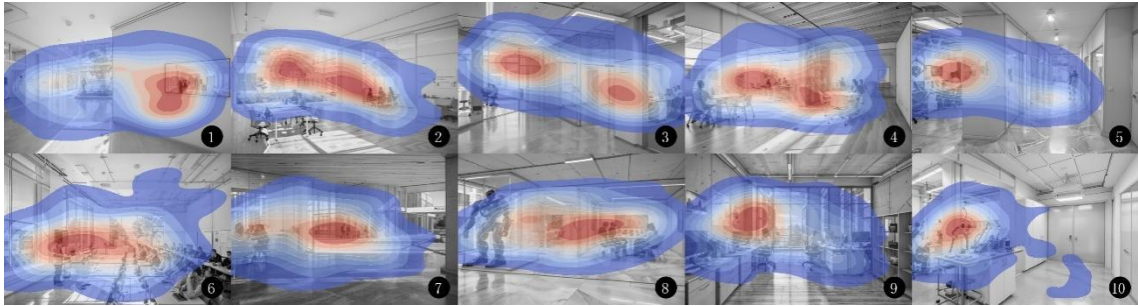


图 6 场景图像的热力图分析 (图片来源: 作者自绘)

### (2) 视觉焦点聚类分析

视觉焦点聚类分析 (Visual Focus Clustering Analysis) 是一种用于理解和分析人们在观看图像、视频或界面时视觉注意力集中区域的方法, 以揭示受试者关注的主要区域。图 7 显示 10 张场景图像的视觉焦点聚类分析结果, 每种颜色代表一位受试者的数据。相较于热力图, 视觉焦点聚类分析能呈现所有注视点的坐标位置, 直观性不足但准确度更高, 若需对兴趣程度进行精准排序, 需要更大的样本量且结合点的概率分布进一步分析量化。

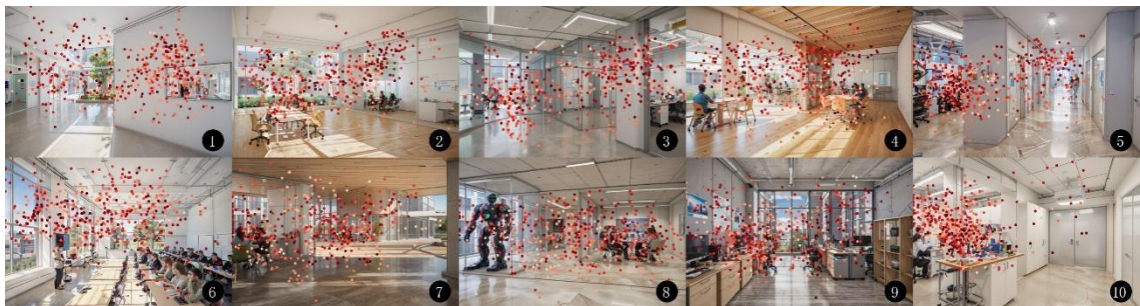


图 7 场景图像的视觉焦点聚类分析 (图片来源: 作者自绘)

### (3) 视线方向分析

视线方向分析 (Gaze Direction Analysis) 是通过跟踪和分析用户的眼动数据, 来理解用户在观看视觉图像时的视线移动路径和方向的研究方法, 能够揭示用户在视觉图像中的注意力分布和信息处理方式, 从而为设计优化提供数据支持。图 8 显示 10 张场景图像的视线方向分析结果, 由浅到深的线段表示观看的时间顺序先后, 多条渐变线段组成一条完整的视线路径图。将所有受试者的视线路径叠加, 浅色线段多的区域代表这个区域容易先被受试者注意到。与统计注视点个数不同视线方向分析可用于设计引导视觉路径, 除互动场景之外, 自然景观、材料色彩等空间要素也能通过优先吸引人的注意, 从而使其周边的信息要素更容易让人获取。

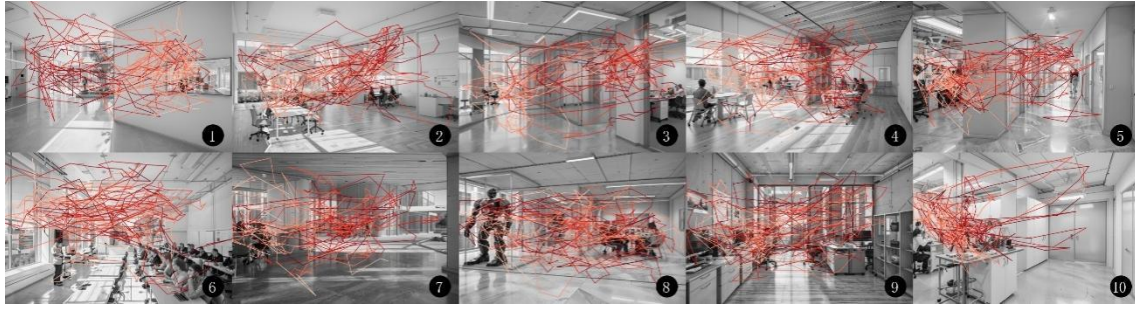


图8 场景图像的视线方向分析（图片来源：作者自绘）

## 4 结果讨论

综合问卷评价和眼动数据采集，可以初步验证科研建筑空间中的信息和空间要素的吸引力程度。信息要素中，互动场景吸引力最强，其次是模型、设备等立体信息媒介，最后是电视、海报、标识等平面信息媒介。互动场景中有人的参与，且包含的信息量最大，建立人与人、人与各个信息媒介的关联性，因此最容易获得关注。模型设备等立体媒介交互性好、体量大，展示的信息更加直观有趣，故也较能吸引人的注意力。电视海报等平面信息媒介展示的信息较为单一，但专业性往往更强，结合立体媒介和互动场景能达到更好的信息传递效果。空间要素中，空间界面透明程度对人的吸引力影响最显著。透明的空间界面能增加空间延伸感，并展现更多所在空间以外的信息，在固定的空间布局中建立不同空间里人与信息的互动关系。此外，界面的材料色彩、自然景观等要素可能有一定视觉引导性，但总体吸引力不强，可结合缓冲和休闲空间设计。

## 结语

研究将网页眼动追踪技术应用到建筑空间视觉图像的分析中，并结合问卷设计、python 可视化分析搭建从数据采集到数据分析的较为完整的工作流。其优势在于，整体研究体现出轻量化、低成本的特点，用一台带摄像头的笔记本电脑即可快速完成数据采集和分析。其不足之处是，WebGazer.js 技术易受到环境干扰，位置和时间的计算不够准确，不适用于需要精细空间解析度的研究，若允许的误差阈值低于 8%，容易出现校准结果不稳定带来的实验结果不准确、效率低等问题。此外，该技术目前仅局限于空间视觉图像的眼动追踪应用，不支持动态场景分析。今后可进一步发掘其集成空间动态场景的可能性，结合视频动画、VR 虚拟现实技术等方法开展实验研究。综上，基于 WebGazer.js 搭建的网页眼动追踪平台适用于低成本且快速地量化的建筑空间视觉吸引力，估算空间兴趣点的大致区域，在建筑相关的科研实验研究、初步调研分析等领域具有一定的应用价值。

## 图表来源

均为作者自绘

## 参考文献

- [1] Higuera-Trujillo J L, Llinares C, Macagno E. The cognitive-emotional design and study of architectural space: A scoping review of neuroarchitecture and its precursor approaches[J]. *Sensors*, 2021, 21(6): 2193.
- [2] Mahamadu A, Prabhakaran A, Clark K, et al. The application of eye-tracking technology in architecture engineering and construction industry: A systematic review[J]. *Industry4.0 Applications for Full Lifecycle Integration of Buildings*, 2021: 56.
- [3] Visual attention predictive model of built colonial heritage based on visual behaviour and



subjective evaluation.

- [4] Maghool S A H, Schnabel M A, Moleta T. A Framework for Quantifying the Temporal Visual Experience of Architecture[J].
- [5] Iñarra Abad S, Vidal F J, Llinares Millón C, et al. VISUAL ATTENTION IN THE EVALUATION OF ARCHITECTURAL SPACES[J]. Revista de EGA, 2015 (25).
- [6] Klatt S, Memmert D. Studying spatial visual attention: the attention-window task as a measurement tool for the shape and maximum spread of the attention window[J]. Frontiers in psychology, 2021, 12: 614077.
- [7] Karmann C, Aydemir B, Chamilothon K, et al. Saliency prediction in 360° architectural scenes: Performance and impact of daylight variations[J]. Journal of Environmental Psychology, 2023, 92: 102110.
- [8] Kim M K. Implementing biosensing based user preference visualisation in architectural spaces[J]. arXiv preprint arXiv:2312.11520, 2023.
- [9] Mahmoud N S, Abdo U A. Using Eye-tracking tools in the visual assessment of architecture[J]. Engineering Research Journal-Faculty of Engineering (Shoubra), 2022, 51(3): 163-174.
- [10] Papoutsaki, A., Sangkloy, P., Laskey, J., et al. Webgazer: Scalable webcam eye tracking using user interactions[C]//Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence-IJCAI 2016, 2016.
- [11] Duchowski A T, Duchowski A T. Eye tracking methodology: Theory and practice[M]. Springer, 2017.
- [12] Kyrk J, Gagas Piechowiak S. Eye-Tracking over Source Code: A Benchmarking Extension to Evaluate the Accuracy of EyesOnTheCode[J]. 2023.
- [13] Erazo M C, Cocha Tobanda E, Yoo S G. Development of a System for Controlling IoT Devices Using Gaze Tracking[C]//International Conference on Intelligent Sustainable Systems. Singapore: Springer Nature Singapore, 2023: 157-171.