

# 基于Unity的智能交互虚拟文物展览馆

夏方方, 郭润甲, 吕镇宇, 刘芳丽, 郭子俊

北京信息科技大学计算机学院, 北京

收稿日期: 2023年10月17日; 录用日期: 2023年11月16日; 发布日期: 2023年11月23日

## 摘要

近年来, 基于虚拟现实技术建设的博物馆建设逐渐成为了全球各大博物馆发展、政府投资、科研开发领域的热点。在数字时代, 将VR技术应用于现代博物馆的建设之中具有重要的现实意义。虚拟文物展览馆作为传统博物馆的延伸改变了传统历史文物的陈列理念, 虚拟博物馆不仅为用户提供了一种访问数字资源的方式, 同时也提供了一种全新的展示和体验文化遗产的形式, 还为文物保护提供了新的思路。本文区别于传统虚拟博物馆的构建方法提出了一种构建可智能交互的虚拟文物展览馆综合系统的方法, 以可智能交互的虚拟文物展览馆作为研究背景, 融入基于AI-driven的虚拟导游、显著性检测、眼动追踪等技术增强与用户的交互性, 为用户带来更丰富的游览体验。同时为国家文化遗产和数字文化发展提供了一种全新的展示和保护手段, 积极的推动文化资源的数字化发展。

## 关键词

Unity 3D, 虚拟文物展览馆, 显著性检测, AI-Driven虚拟导游, 眼动追踪

# Virtual Cultural Relics Exhibition Hall Based on Unity Intelligent Interaction

Fangfang Xia, Runjia Guo, Zhenyu Lv, Fangli Liu, Zijun Guo

Computer School, Beijing Information Science and Technology University, Beijing

Received: Oct. 17<sup>th</sup>, 2023; accepted: Nov. 16<sup>th</sup>, 2023; published: Nov. 23<sup>rd</sup>, 2023

## Abstract

In recent years, the construction of museums based on virtual reality technology has gradually become a hot spot in the fields of development, government investment, and scientific research and development of major museums around the world. In the digital age, the application of VR technology to the construction of modern museums has important practical significance. As an extension of traditional museums, virtual cultural relics exhibition halls have changed the display

concept of traditional historical cultural relics. Virtual museums not only provide users with a way to access digital resources, but also provide a brand-new form of displaying and experiencing cultural heritage, and also provide new ideas for cultural relics protection. Different from the construction method of traditional virtual museums, this paper proposes a method for constructing a comprehensive system of virtual cultural relics exhibition halls that can be intelligently interacted. The virtual cultural relics exhibition hall that can be intelligently interacted is used as the research background, and technologies such as AI-driven virtual tour guides, saliency detection, and eye tracking are integrated to enhance the interaction with users, bringing users a richer tour experience. At the same time, it provides a brand-new means of display and protection for the development of national cultural heritage and digital culture, and actively promotes the digital development of cultural resources.

## Keywords

Unity 3D, Virtual Cultural Relics Exhibition Hall, Saliency Object Detection, AI-Driven Virtual Tour Guide, Eye Tracking

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

### 1.1. 国内外研究现状

虚拟文物展览馆方面：在新世纪我国十分注重博物馆的建设与发展，对虚拟文物展览馆建设也进行了深入研究。在技术的支持下，我国研发了虚拟文物展览馆的新兴产物，如腾讯博物馆。腾讯博物馆具有线上展览、线下导览等多重功能，为故宫博物院、法国国家博物馆联合会等组织提供了新的思路。但相比于西方发达国家比较成熟的文化虚拟博物馆的构建技术[1]，我国虚拟文物展览馆建设仍处于起步阶段，应继续加大探索力度。基于国内现有的虚拟博物馆构造方法和经验[2]，我们要考虑构建交互性更强的智能交互虚拟文物展览馆。具体来说在构建虚拟文物展览馆时我们积极探索了如何充分发挥各种技术手段在展览中的作用，如将单调的固定路线游览升级成为可自由移动的参观游览、基于用户意图识别的自动触发展品文物视频讲解。

眼动追踪方面：近年来，随着计算机视觉、图形学、人工智能等新兴技术的发展及应用需求的增加，“眼动 + VR”正在被越来越多的大企业所关注。苹果、谷歌、微软、Magic Leap、三星等多家公司都已布局眼动追踪技术相关业务。目前的眼球追踪、视线估计技术比较完备，但是多用于游戏和医疗领域，目前尚无应用在虚拟文物展览馆的场景构建中。

显著性目标检测方面：随着计算机视觉和图像处理技术的发展，显著性检测成为一个重要的研究领域。过去的方法主要基于低级特征(如颜色、纹理和边缘等)进行显著性分析。而近年来，卷积神经网络、Attention 等技术被应用到了显著性目标检测中。当前的显著性检测算法主要可分为以下几类：基于聚合的模型、基于边缘的模型、基于 Vision Transformer 的模型。基于聚合的模型通过提取多个维度的特征，这些特征经过一些网络层的处理后再融合成为最终的显著性图；基于边缘的模型通过以边缘和掩码为监督，来提高算法的性能；基于 Vision Transformer 的模型引入 Vision Transformer 来更好地获取图像特征。目前，显著性目标检测技术应用在虚拟文物展览馆的实例较少。

数字虚拟人方面：自 2017 年开始数字虚拟人技术进入快速发展阶段，二次元圈层用户崛起，人工智能、机器学习、渲染、VR 等技术快速兴起，短视频等多种视频社交形式加速普及，产业资本大规模进入该领域，大量数字虚拟人被开发应用，虚拟偶像、虚拟主播、虚拟主持人等多种形象开始出现。但目前虚拟导游多采用动作捕捉技术生成，如国家博物馆的虚拟数智人“艾雯雯”和“全古今”，但是存在虚拟人建模所需设备昂贵，数据采集过程繁琐复杂的问题[3]。本系统采用 AI-driven 的虚拟人作为虚拟导游，避免了复杂的数据采集与建模步骤。具体来说，首先基于 Stable Diffusion 模型生成虚拟人物形象，然后利用语音合成技术将文本数据转换为音频数据，再基于说话人脸生成技术(TFG)生成嘴部动作与音频数据一致的说话人脸视频，最后基于 GFPGAN 超分辨率模型增强面部清晰度生成具有高分辨率的数字虚拟人。

## 1.2. 研究意义

研究构建可智能交互的虚拟文物展览馆具有以下重要意义：

对于用户来说，可以提供一种交互性更强的参观博物馆的方式。用户可以不受空间和时间限制，真实的感受历史场景，能提升人们对文物历史和文化遗产的认知，增强文化素养和自信。

与传统实体博物馆相比，此技术不受保护、修复和复原条件的限制，可以将博物馆真实完整地存储到计算机网络中，实现三维数字存档，为国家文化遗产和数字文化发展提供了一种全新的展示和保护手段，推动文化资源的数字化发展。

与传统虚拟博物馆相比，此方法为如何提升虚拟博物馆的智能性和交互性提供了新的思路，虚拟博物馆可以融合多种交互技术为用户提供多感官、多方式的体验。

本文提出一种基于 Unity 3D 构建互动的虚拟文物展示馆的方法，用于将实体博物馆数字化展示，为用户提供数字化展览。

## 2. 开发环境介绍

### 2.1. 主要特征

在进行虚拟文物展览馆系统开发时，主要在 Unity 3D 引擎中采用 VR 技术开发。Unity 3D 是一款在游戏开发和虚拟现实领域广泛应用的跨平台游戏引擎。它融合了多项独特特性，为开发者提供了强大的工具和极大的灵活性，能够打造出引人入胜的游戏体。

Unity 3D 的主要特征包括跨平台支持、强大的图形引擎、物理引擎、脚本语言、资源管理、社区和生态系统、实时预览、VR 和 AR 支持。

在进行 AI 虚拟人生成、眼动追踪任务时，主要使用了深度学习技术。在 AI 虚拟人生成中，深度学习技术可以实现对人物形态、面部表情和语音等方面的智能建模和仿真，从而实现更加真实和自然的虚拟人形象。在眼动追踪中，深度学习技术可以实现对眼部运动的实时追踪和预测，从而更加准确地解析用户的视线和注视点。

### 2.2. 关键技术概述

Unity 3D 支持多个虚拟现实设备，如 HTC Vive、Oculus Rift 和 PlayStation VR，并提供了丰富的交互控制选项，包括虚拟手柄、手势识别和触摸控制等技术。它还支持头部追踪、空间音效和高质量图形渲染等技术，以提高虚拟环境的真实感和沉浸感，同时注重性能优化，确保平滑流畅的运行。

深度学习技术是一种基于人工神经网络的机器学习方法，它模仿人脑神经元之间的联结方式和信息传递过程，通过多层次的神经网络来学习和表示数据的复杂模式。深度学习技术的核心是深度神经网络(Deep Neural Network, DNN)，它由多个网络层组成，每一层都包含大量的神经元。不同层之间的神经元

通过权重连接, 并通过激活函数来实现非线性映射, 从而实现对数据的分层抽象和特征提取。

### 3. 虚拟文物展览馆的构建步骤(图 1)

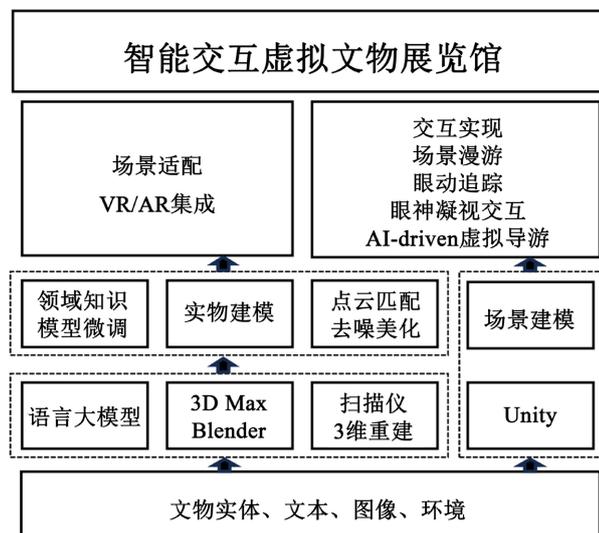


Figure 1. Architecture diagram of virtual museum

图 1. 虚拟文物展览馆整体架构图

#### 3.1. 项目规划

确定建设虚拟文物展览馆的目标和内容, 并选择适合的技术平台进行开发。收集博物馆展品的相关资料并整理分类, 确保版权和授权问题。设计虚拟文物展览馆的场景布局、展品展示方式和用户交互方式。

#### 3.2. 三维模型的创建

使用 Blender 等建模软件、使用扫描仪对文物扫描和深度学习三维重建的方法对文物进行三维模型的创建。利用建模工具创建各种复杂的三维模型, 并进行编辑和变形以调整模型的形状、细节和纹理等方面。添加材质和纹理来模拟不同材质的外观和光照反射效果(图 2)。

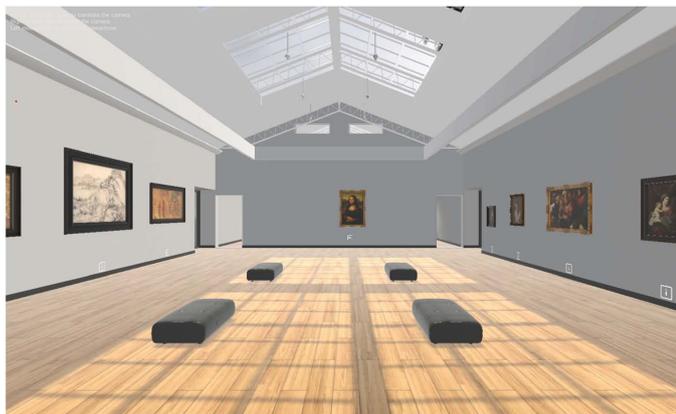


Figure 2. Rendering of the virtual museum

图 2. 虚拟文物展览馆效果图

### 3.3. 互动功能的实现

互动功能主要分为基础功能和高级智能交互功能两部分。基础功能部分实现简单的事件触发。在 Unity 3D 中,为实现与文物进行互动,采取对物体绑定 Interactable 组件和 Box Collider 组件添加感应区域来实现感应手部控制器对物体的接触,添加 Throwable 组件来实现对物体的抓取和释放的机制,同时为了让物体释放后有更加现实的物理特性,添加 Rigidbody 组件,最终实现对物体的互动。

### 3.4. 动态导览系统设计

使用 Unity 的场景编辑器创建导览系统的场景。根据需要添加建筑物、地形、道路等元素,使场景准确反映实际环境。配置主相机以使用户可以自由浏览场景,并设计用户交互界面,如按钮、面板、信息提示等。使用导航系统创建路径,帮助用户沿着指定路径移动并提供导航指示。关联信息与关键元素,将文字、图片、视频等与场景中的元素关联起来,在用户导览时提供相关信息。

## 4. 智能交互技术的应用

为了提升用户参观虚拟文物展览馆过程中的互动性,系统采用 AI-driven 虚拟导游、凝视关注、眼动追踪技术(图 3)。



Figure 3. Virtual tour to explain cultural relics  
图 3. 虚拟导览讲解文物资料

### 4.1. AI-Driven 虚拟导游

传统虚拟博物馆往往没有选择引入虚拟导游来增强游客的体验感[4],仅通过视觉浏览虚拟文物展览馆无法取得交互性更强的体验,我们制作并引入了虚拟导游进行文物讲解。首先我们从互联网收集到博物馆中文物展品的相关信息并进行清洗和整理。利用 tts (语音合成)技术将文本数据转换成音频数据,为基于语音驱动的虚拟导游提供了声音素材。由于直接使用现实中人物形象作为虚拟导游会涉及版权的原因,我们采用 stable diffusion 的图生图模型生成并保存虚拟人物形象。结合已经处理好的声音素材,我们借助 TFG (Talking Face Generation)技术生成讲解视频,由于大部分 TFG 模型是基于英文数据集训练产生,对中文数据处理效果不佳[5],于是我们采用了基于二维说话人生成模型 SadTalker [6]驱动 Stable Diffusion 生成的虚拟人物形象制作虚拟人讲话视频。但制作的视频清晰度往往不能达到理想水平,因此我们对比多种高分辨率模型最终采取 GFPGAN [7]生成网络来进行虚拟人脸的高分辨率处理,进行了脸部增强,提高清晰度的同时也使说话人脸更加逼真。经过上述流程最终生成能够流畅处理中文语音的高分辨率虚拟导游形象,制作成视频格式导入 Unity 中的虚拟博物馆项目中(图 4)。

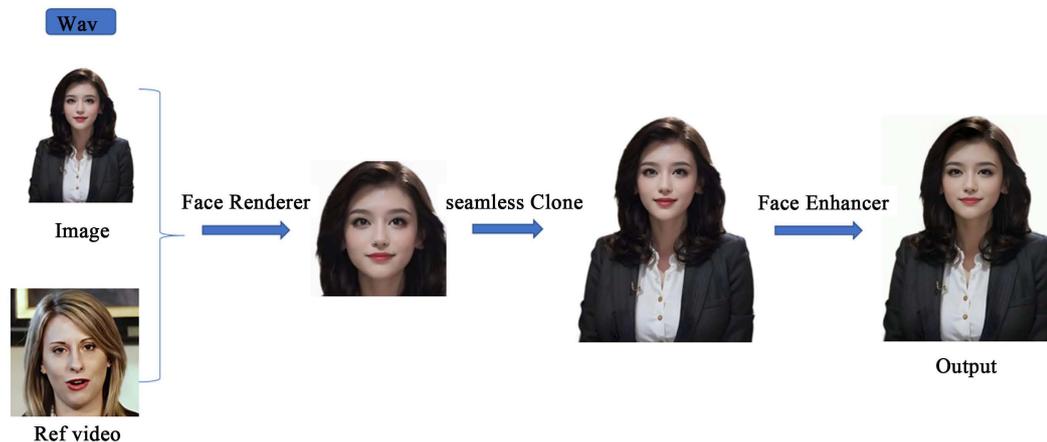


Figure 4. Flowchart of digital image generation  
图 4. 数字形象生成流程

## 4.2. VR 场景中的眼动追踪与凝视关注

眼动追踪，是指用户在虚拟文物展览馆应用场景之中，对 UI 交互菜单显示界面中的任务模块通过凝视操作进行选择，触发对应的事件，实现相应的模块功能。眼动追踪功能主要通过 SteamVR Plugin 框架中的头盔显示器凝视脚本 SteamVR GazeTracker.cs，此脚本为凝视功能的开发提供了接口，同时依托于 Unity 3D 中的射线碰撞检测机制得以实现。

眼动追踪的应用能够使视觉交互更加自然，提升沉浸感和逼真度，同时也能增强交互的精确性和灵活性(图 5)。

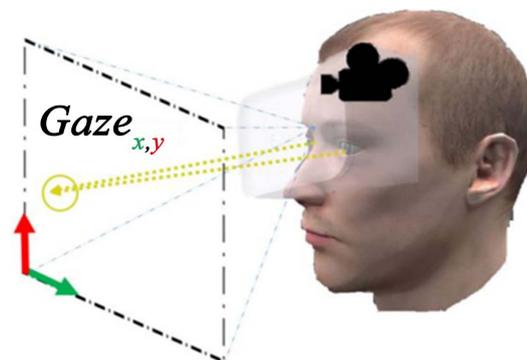


Figure 5. Schematic diagram of eye tracking technology  
图 5. 眼动追踪技术示意图

我们通过对用户的凝视关注区域进行显著性目标检测来间接实现眼动追踪。显著性目标检测的目的是在图像或视频中准确地检测出人类视觉感知上更显著的、具有显著性的目标物体。将显著性目标检测应用到虚拟文物展览馆中，可以引导参观者注意，从而提高其浏览和体验效。4.2 中所提到的虚拟导游部分也可根据显著性目标检测的结果，来针对这些区域提供导览和解说服务，帮助参观者更好地理解 and 欣赏展品。通过显著性检测，也可以确定艺术品中重要和突出的部分，可以评估展览的效果和吸引力，从而帮助博物馆设计者更好地对虚拟文物展览馆进行展览规划和改进。

针对 VR 文物博物馆的特殊场景，为了更好的应用显著性目标检测算法，我们参考先进的深度学习算法 LDF [8]的网络结构并对其进行改进；以 vit-base8 为主干网络进行特征提取，将提取到的特征向量经过

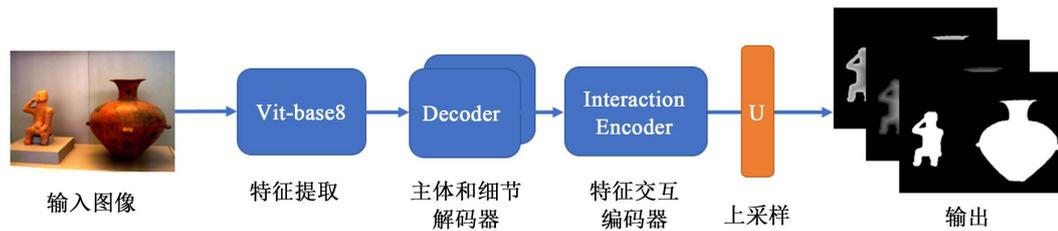
形状变换后输入到两个解码器中，解码器的输出结果再进行特征交互编码器。特征交互编码器的输出再次输入到两个解码器中进行信息交换，最后得到显著性图，其大致网络结构如图 6 所示。显著性目标检测的效果如图 7 所示，其中 Image 表示输入图像，Pred 表示预测结果。

我们用开放数据集 DUTS 来训练我们的模型，并在 DUTS-T、ECSSD、HKU-IS 三个数据集上进行评估。并与 LDF、M3NET [9]、VST [10]等方法进行对比，分别对比了平均绝对误差(MAE)、平均 F 测量(mF)、E 测量( $E_{\xi}$ )。对比结果表 1 所示，从表中可以看出，我们的网络性能略优于其他方法。

**Table 1.** Comparison results of salient object detection with advanced methods

**表 1.** 显著性目标检测与先进方法对比结果

Algorithm	ECSSD			HKU-IS			DUTS-TE		
	1000 images			447 images			5019 images		
	MAE	$mF$	$E_{\xi}$	MAE	$mF$	$E_{\xi}$	MAE	$mF$	$E_{\xi}$
VST	0.033	0.920	0.918	0.029	0.900	0.953	0.037	0.818	0.892
LDF	0.034	0.930	0.925	0.028	0.914	0.954	0.034	0.855	0.910
M3NET	0.029	0.926	0.918	0.026	0.919	0.958	0.037	0.855	0.906
ours	0.031	0.930	0.927	0.027	0.921	0.956	0.031	0.864	0.916



**Figure 6.** The framework of saliency object detection module

**图 6.** 显著性目标检测的网络结构



**Figure 7.** Saliency object detection renderings in the scene of cultural relics museum

**图 7.** 文物博物馆场景下显著性目标检测效果图

## 5. 虚拟文物展览馆的效果与优势

虚拟文物展览馆的效果和反馈非常好，观众不再受制于时间和空间的限制，可以根据自己的时间安排自由参观。同时眼动追踪能获取观众的意图，显著性目标检测能锁定观众的兴趣范围，增强了交互的便捷性。同时，虚拟导游的引入可以为用户带来听觉上的体验，使在参观虚拟文物展览馆时在更多感官上获得极致体验。虚拟文物展览馆的推出也获得了专家和学者的认可与支持，将为学术研究和文化交流

提供了新的平台和途径。

### 5.1. AI-Driven 虚拟导游讲解

虚拟导游是虚拟文物展览馆中一项引人注目的创新技术。通过 AI 技术的应用,虚拟导游不仅可以引领参观虚拟文物展览馆并且进行讲解,还能根据参观者的需求和兴趣,为他们量身定制导览内容。这种个性化的导览服务使得参观者能够更加深入地了解 and 欣赏所展示的内容,提升了参观体验的质量。虚拟导游的引入使得虚拟文物展览馆得以吸引更多喜欢和想体验虚拟现实的人涌入虚拟文物展览馆。

### 5.2. 即时访问

在使用 Unity、Blender 等软件进行三维模型的创建过程中,虚拟文物展览馆通过运用先进的技术手段,如虚拟现实(VR)、增强现实(AR)和全景摄影等,为观众营造了身临其境的参观体验,观众可以通过佩戴 VR 头显或使用智能设备,自由选择参观路线进行探索展厅、观看文物、欣赏艺术品等活动。

同时虚拟文物展览馆具有无时无刻的访问的优势,观众可以根据自己的时间安排自由参观,不必为时间与地理位置而感到局限和束缚。特别是对于那些生活在偏远地区或者行动不便的人们来说,虚拟文物展览馆为他们提供了更为便捷的途径来欣赏艺术品与文化遗产[3]。虚拟文物展览馆即时访问的优势可以让我们以更低的成本去随时访问虚拟文物展览馆,让博物馆以数字虚拟的形式永远保存在互联网中。

### 5.3. 注意力引导

眼动追踪与凝视关注模块可以检测出用户感兴趣的区域。在文物展览馆中,当用户视线停留在某一区域超过 2 秒,系统会将该区域确定为用户感兴趣的区域,并弹出该区域内所展示文物的相关介绍,同时虚拟导游也会为参观者提供精确的面部动画和语音讲解。对于博物馆布局和建设者而言,可以利用这些用户感兴趣的区域来为展览馆进行更好地布局。

## 6. 展望与总结

**渲染技术提升:**随着虚拟现实技术不断发展,实时渲染技术的不断改进,虚拟文物展览馆中的文物将以更高质量的图形效果展现在观众面前,同时会支持多种平台访问游览,包括 PC、移动设备、游戏主机和云平台等。

**社交互动功能增强:**基于虚拟现实技术的虚拟文物展览馆将会着重提升用户互动体验,将来用户可以在虚拟空间中与他人分享参观体验。

**融合大语言模型的虚拟导游:**未来虚拟文物展览馆中的虚拟导游将会与大语言模型融合,能够即时响应用户的需求,为用户带来生动、准确、有趣的智能导览。

随着技术的进步,基于虚拟现实技术构建的虚拟文物展览馆有望继续演进和创新,将会提供更加真实的体验和更加智能化的功能。虚拟文物展览馆借助社交媒体等渠道与用户进行互动和交流,引导用户进行更加深入的探究和发现,还可以与其他数字文化遗产展示和教育工具相结合,创新展示形式和教育方式,为用户提供更加丰富、多样的文化体验,具有广阔的发展前景和应用空间。未来,虚拟文物展览馆将继续演进和创新,为用户带来更加真实、智能化、个性化的参观体验,同时也将推动文化遗产的传承和推广。

## 基金项目

由北京信息科技大学促进高校分类发展大学生创新创业训练计划项目——计算机学院(5112310855)支持。

## 参考文献

- [1] Eriksson, J., Widen, P., Ferrari, S. and Hayashi, M. (2023) Digital Reconstructions of Picture Galleries as Art Historical Method: The Virtual Museum at the Royal Palace in Stockholm. *Digital Applications in Archaeology and Cultural Heritage*, **29**, Article e00268. <https://doi.org/10.1016/j.daach.2023.e00268>
- [2] Mu, Z.H., Liu, M. and Tan, Z.X. (2018) On the Interactive Display of Virtual Aerospace Museum Based on Virtual Reality. 2018 *International Conference on Robots & Intelligent System (ICRIS)*, Changsha, 26-27 May 2018, 144-146. <https://doi.org/10.1109/ICRIS.2018.00045>
- [3] 许诺. 虚拟时代“文旅元宇宙”的变现探索[J]. *数据*, 2022(9): 26-28.
- [4] 王经田, 梁晓琳, 王志飞. 基于 VR 技术的虚拟博物馆设计与开发[J]. *现代信息技术*, 2023, 7(15): 29-34.
- [5] 宋昕洋, 阎志远, 孙沐毅, 戴琳琳, 李琦, 孙哲南. 说话人生成研究现状与发展趋势[J]. *计算机科学*, 2023, 50(8): 68-78.
- [6] Zhang, W.X., Cun, X.D., Wang, X. and Zhang, Y. (2023) SadTalker: Learning Realistic 3D Motion Coefficients for Stylized Audio-Driven Single Image Talking Face Animation. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, 17-24 June 2023, 8652-8661. <https://doi.org/10.1109/CVPR52729.2023.00836>
- [7] Wang, X.T., Li, Y., Zhang, H.L. and Shan, Y. (2021) Towards Real-World Blind Face Restoration with Generative Facial Prior. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 9164-9174. <https://doi.org/10.1109/CVPR46437.2021.00905>
- [8] Wei, J., Wang, S.H., Wu, Z., Su, C., Huang, Q.M. and Tian, Q. (2020) Label Decoupling Framework for Salient Object Detection. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 13022-13031. <https://doi.org/10.1109/CVPR42600.2020.01304>
- [9] Yuan, Y., Gao, P. and Tan, X.Y. (2023) M3Net: Multilevel, Mixed and Multistage Attention Network for Salient Object Detection. ArXiv: abs/2309.08365(2023):n.pag.
- [10] Liu, N., Zhang, N., Wan, K.Y., Shao, L. and Han, J.W. (2021) Visual Saliency Transformer. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 10-17 October 2021, 4702-4712. <https://doi.org/10.1109/ICCV48922.2021.00468>