Artificial Intelligence and the Arts:
Towards AI-Guided Accessible Learning Spaces in Virtual Reality

A.Weissman, Lucas – Bouabid, Marwa – Stevens, Logan – Steigelman, Joey

Introduction

This research project aims to identify how AI and XR can be used to promote accessibility in learning spaces, and what techniques can improve user perception. First, the project captured details of one's real-life experience in a museum environment, and how senses such as vision, touch, and hearing could be emulated in a virtual space. Then, we researched assistive technologies found in AI (e.g., Image Captioning) and VR (e.g., increased object interactivity, magnification, and color manipulation). Combined, these can enhance perception in a Virtual Environment (VE).

Methodology

In order to increase access for users, we aimed to develop integrated AI captioning, in addition to our own set of assistive VR tools, in order to help users accurately perceive the VR scene and the artworks within it, regardless of ability.

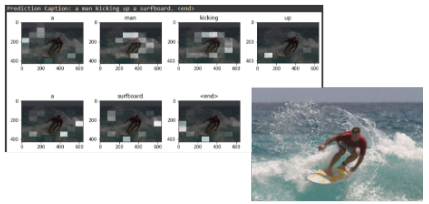
Findings

We observed that our AI-generated image captions were of inconsistent and often insufficient quality. Realtime AI caption generation was also impractically slow due to cloud-run AI systems.

Acknowledgments

We thank the following groups/individuals for their contributions:

Brown University, Google Research, Dr. James Tompkin, Catherine Chan, Ji Won Chung, and Qian Zhang



The **MS-COCO** dataset was used to train our AI model for image-captioning.

"a man riding on a sail"

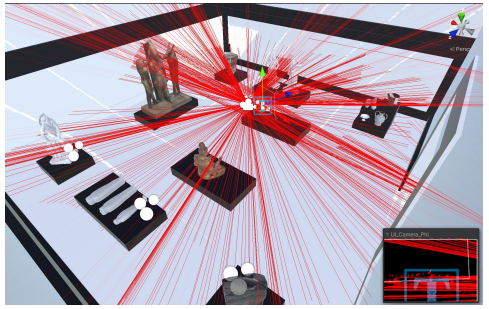
"picture of ladies standing next to the window"

"a woman sitting in front of a cushion on her phone"

Implementation

In the Unity Engine, a VE was created replicating a museum in which users could interact with the artworks in order to summon AI-generated captions to aid in experiencing the visual art. Our goal was to link an API to Unity such that it would take the current piece of art that the user is seeing, and externally run a dedicated AI on that artwork to generate captions in real-time.

Demonstration of real-time AI-captioning within the VE



Raycasting elements via multiple cameras to keep track of what the user is seeing

Detailed paintings and 3D artifacts were sourced and placed within the VE. Next, we were able to observe which artwork was being viewed by the user via raycasting. Then, the visual data of that artwork was sent to our AI and analyzed to generate a brief description (caption) of the artwork. Features such as these improve accessibility of artworks for individuals of all abilities.

Conclusions and Future Steps

- With increasingly efficient AI algorithms, specific applicable datasets, and/or more advanced computer hardware, accurate live AI caption generation would be more practical.
- Training the AI on a larger and/or more applicable dataset will improve the quality of the captions.
- Rebuilding and implementing elements from the depreciated SeeingVR Toolset (e.g., magnification, object-outlines, text-to-speech, etc.) will further improve accessibility of the VE and artworks.
- User studies would be beneficial in better adapting such toolsets and VEs to users of all abilities.