LAVIN: An AI-Navigated Art Experience in Virtual Reality

Weidi Zhang, Jieliang Luo

Arizona State University, Autodesk Research Arizona, USA; California, USA
zhangweidilydia@gmail.com, rodgerljl@msn.com

Abstract

This paper outlines the conceptual background, design methodology, and future directions of *LAVIN*. This virtual reality (VR) artwork provides an immersive experience to visually explore one understanding of a neural network in which the real-world maps to 50 daily objects. In the art installation, the neural network constantly analyzes the surrounding environments via a camera and outputs real-time semantic interpretations, which navigate the audience in a virtual world consisting of all the fluid abstract structures of daily objects that the neural network can recognize. We create these fluid virtual structures using data visualization, photogrammetry, and 3D modeling. By merging Artificial Intelligent (AI) system design with VR world-building, *LAVIN* offers an immersive art experience for symbiotic imaginations that questions the values and beliefs in the modern AI age.

Keywords

AI System Design, Virtual Reality, Immersive Media Design, Generative Art, Interactive Art Experience.

DOI

Introduction

In the last decade, object detection has made tremendous progress. In some narrow domains, the algorithm has almost reached human-level accuracy [1]. Each neural network's output is a unique projection of its understanding of the real world, regardless of whether it has been trained to recognize thousands of objects or only produces binary outputs [2]. Regardless of the complexity of the projection, it shapes the "world value" of the neural network [2]. Numerous deep neural networks are implemented for various intelligent applications. Despite this, even the most complex neural network model is not capable of representing the real world entirely. A neural network used for facial recognition, for example, can only interpret the input as a collection of faces [2]. Thus, our design question arises: what is the ground truth in the modern AI age?

As a conceptual response to the question regarding the worldview and value of neural networks, *LAVIN* provides one understanding of a neural network by combining VR worldbuilding and AI system design. We design our AI system to observe surrounding environments in the gallery space via a camera and output one word to describe the live-streaming based on the neural network's interpretation. The output word is the navigation information that travels the participants in the VR in real time. The AI system is trained to only understand 50 daily objects which leads to its misinterpretation of the live-streaming information. The VR world is made up of these 50 objects our AI system can recognize. *LAVIN* aims to provide an immersive art experience that is navigated by an AI system through visualizing the imagined world of a neural network artistically.

Figure 1. *LAVIN*, Virtual Reality Art Installation, Siggraph Art Gallery at LA convention center, ©Weidi Zhang.

Background

The real world is a sophisticated collage with diverse groups of creatures, overloaded data, a myriad of choices and chances, fleeting messages, complex shifting meanings, and distinct stories, cultures, and political systems. If we perceive the images of our world as a visual representation that is programmable, networked, and operative, what is the method to associate different visual information? How do we use an algorithmic approach to create a new kind of chance and choice operations in the art-making process? Who will be the author: machines, humans, or human-machine hybrid? What are the data-driven material, forms, and aesthetics? How will the audiences interact with this assemblage, and in which way will their decisions change the experience?

Seymour Chatman states that: In this age of mechanical and electronic production and reproduction, it would be naive to reject the notion of nonhuman narrative agency, for instance, a story constructed through algorithmic processes or as a result of multiple voices constructing together in real-time [3]. With the goal of integrating a non-human narrator for immersive storytelling, we designed an AI system that navigates this VR journey based on its real-time understanding of the surrounding environment. The customized machine reinterprets the real world to build topographical associations between objects in a VR journey.

Related Artworks Worldview of A Neural Network

Interactive works employing machine learning algorithms often offer a user-centric and adaptive experience. Using autonomous systems, responsive applications can assemble artifacts, discoveries, traces, and experiences based on the prior knowledge of the machines and their interpretations of audiences' inputs.

New media artist Memo Akten's series of interactive works—*Learning to see*—"use state-of-the-art machine learning algorithms to reflect on ourselves and how we make sense of the world" [4]. In the art installation, Memo uses a live camera to capture everyday objects and uses a number of neural networks to analyze the live-streaming feed. The audience is invited to move the objects under the camera and see the visualization evolve based on the changing composition of the moving objects. The visualization changes every 30 seconds to different networks trained on different datasets: water, air, earth, and cosmos. Memo Akten

considers the way of seeing in the conscious mind as a reconstruction based on expectations and beliefs instead of a mirror image of the outside world [4].

Similarly to Memo Akten's *Learning to See*, *LAVIN* emphasizes the importance of machines' prior knowledge and their worldviews in presenting how machines interpret the world using an artistic approach. *LAVIN* differs from the relevant artworks in that it does not directly display machine interpretation as visual output, but rather uses it as navigational information for engaging the author's artistic imagination of a neural network's mind. *LAVIN* combines AI's decision with artistic originality to invent a virtual world of symbiotic imagination.

Navigation in VR Journey

Traditionally, immersants (the participants who wear the head-mounted device) can use VR controllers to navigate a virtual space, which often requires a learning curve. At the same time, there are many VR artworks that investigate conceptual and novel navigation methods.

For example, the pioneer VR artwork, *Osmose*, created by Char Davis in 1995, uses bio-data of breathing to transport users to different elevation places within a virtual environment. *Osmose* presents a multi-sensory immersive experience that engages visitors in a virtual world made of simulation of nature scenery and text. Immersants can navigate themselves in the virtual space through body breathing—a chest-hugging rubber vest that contains sensory devices sensitive to the body's breath—a natural interface [5]. This interface works similarly to diving—when you inhale, you rise. *Osmose* juxtaposes the insecurity of body presence and an emotional sense of tranquility.

Osmose precurses to the booming industry of VR as a platform of narrativity and entertainment in the 21st century. It also challenges the way we view pictorial artworks. The infinite poetic space is unfolding like a canvas waiting for artists to infuse new ideas. A 3-dimensional space adds one dimension to the 2-dimensional pictorial work for the multi-media assemblage. The exploration of the relation between 2-d elements changed into the examination of neighboring topological notions within a dynamic spatial volume.

Similarly to *Osmose* seeking an innovative way of VR navigation for artistic purposes, *LAVIN* creates an intelligent system to contribute to the dynamic flow of the spatial assemblage. Real-world images are connected with virtual space for a unique way of seeing and controlling.

Design Methodology of *LAVIN* System Design

LAVIN provides a conceptual and artistic response to the questions of the Ground Truth in the modern age of AI by designing the navigation system based on a live neural network's real-time interpretation of live streaming data of the surroundings.
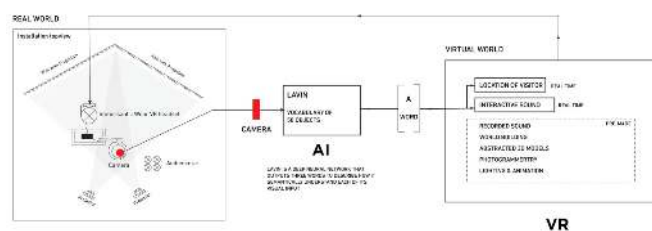

Figure 2. System Map of *LAVIN*. ©Weidi Zhang.

For our AI system design, we specifically use ResNet50, "a pre-trained deep neural network that can be altered for different use cases" [6]. ResNet 50 networks have been trained on the large datasets from Imagenet and "modify the last layers to quickly produce models to tackle new problems" [7]. We build a custom dataset of fifty daily objects and trained a neural network to only recognize the selected fifty daily objects. When an image contains objects outside *LAVIN*'s vocabulary, its limited training dataset leads to a possible false interpretation of the given image.

In the art installation (see Figure.2), *LAVIN* constantly chooses one object to describe live streaming data captured by a camera (set in the center of the exhibiting space). The real-time one-word description travels an immersant to the related objects in the virtual space. Our intelligent system (*LAVIN*) triggers absurd but interesting results, which transports the VR immersant to experience how a limited machine perceives reality.

VR Worldbuilding

The immersive virtual space is a spatial assemblage and an artistic imagination of the neural network. *LAVIN* incorporates both real-world data and artistic manipulation to present a symbiotic imagination. This

multi-layered, bizarre, and unknown VR space is mainly designed by addressing the following three aspects: 3d reconstruction, texture development, and spatial design.

The virtual reality environment consists of fifty objects that the trained neural network (*LAVIN*) can recognize. These virtual sculptural objects are constructed by using the photogrammetry technique—photographs are taken volumetrically of these 50 objects, which are then analyzed, calibrated, and calculated to generate the XYZ position data of the point cloud for the 3d reconstruction (see Figure 4).

Figure 3. Screenshot of the VR world in *LAVIN* ©Weidi Zhang.

Figure 4. Screenshot of the Virtual Reality World in *LAVIN* ©Weidi Zhang.

The aesthetic outcomes of the reconstructions are realistic with a great many unnecessary details for artistic visual representation. We simplify, filter, and rearrange the dataset to distort the constructions and sculpt the shape, at the same time still maintaining the characteristics to make sure the forms of these objects are recognizable (see Figure 5).
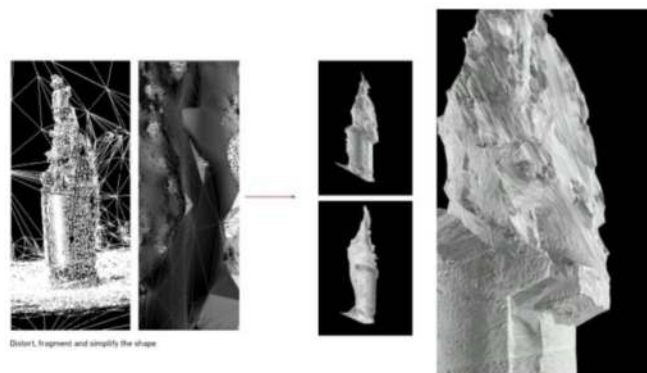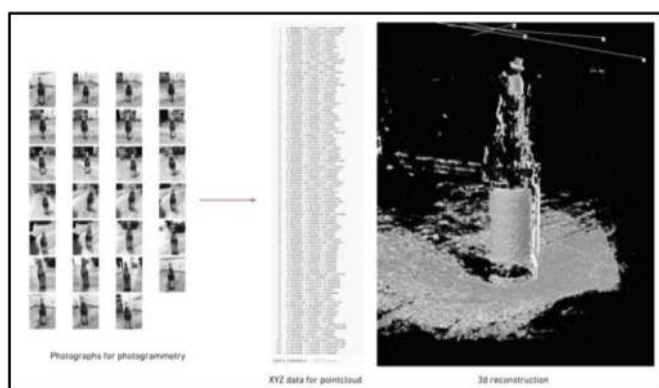

Figure 5. The Sculptural Virtual Objects of *LAVIN*. ©Weidi Zhang.

We keep all fifty 3d constructed forms as sculptures with plaster material in VR. Instead of directly attaching photographs of fifty objects as textures to the 3d forms, we process images as interactive patterns floating in the virtual space, which moves in distinct directions and speeds based on audiences' positions. The pictorial crossovers generate unexpected and poetic meanings between unrelated objects (see Figure 6). In the virtual space, immersants are revolved around the moving textures, witnessing the layers of shapes weave the artistic imagination and abstraction of the neural network.
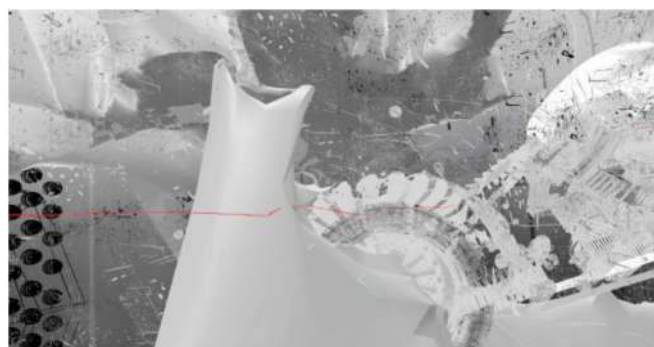

Figure 6. Screenshot of the VR world in *LAVIN* ©Weidi Zhang.

The assembled 2d pictorial artworks often navigate viewers' gazes in the composition by drawing their attention to the correlations between unrelated elements in various materials and shapes. Correspondingly, the navigation of an assembled VR space examines neighboring topographical notions to convey meanings.

The spatial configuration of *LAVIN* is inspired by Paul Klee's Fish Magic. "He creates a magical realm where the aquatic, the celestial, and the earthly intermingle" [8]. The mysterious and sophisticated space on the canvas where fish and flora float among human beings and clock towers. The objects are in indeterminate scale, scope, and direction.

Impacted by Klee's creative solution for imaginative space, *LAVIN* completely abandons the real-world objects' proportion, scale, and properties to emphasize the unpredictability of a wholly imagined world (see Figure 7). The fifty sculptural forms of objects are static like the artifacts displayed in the museums, while the interactively moving patterns densely and dynamically fill the void between them. Metaphors are born out of the associations between texture layers and sculptural forms. By gazing at the textures of the patterns closely, participants can root out the real-world evidence of the objects.

Figure 7. Screenshot of the VR world in *LAVIN* ©Weidi Zhang.

Sound Design and Art Installation

A crucial part of the design of this immersive experience is the use of sound. Inspired by musique concrète, using field recordings, instrumental sounds, and human voices as sound materials, We organize and process them to create sound objects. We then attach them to various static or moving virtual objects (models and patterns) to produce a phantasmagorical soundscape.

Specifically, we record Google Translate's generic human voice reading the words of the 50 objects and attach them to the corresponding objects in the virtual space. Field recordings of cityscape were made as the raw sound materials which are sorted and organized by pruning and splicing, then transformed by using software (Ableton Live and Adobe Audition) to filter and reverberate. The processed sound objects are imported into the virtual world and attached to the different floating textures. When multiple channels of sound cross over and assemble in space, a dynamic spatial montage is created.

This work was presented in the Siggraph Art Gallery (see Figure 8) and SwissNex Gallery in 2019. The exhibition design includes a screen or two projections, a webcam, a VR headset, a speaker, and a PC. A camera is set in the center of the space to observe surroundings and send live streaming to the system. The VR headset is placed on top of a pedestal in the space. The real-time VR experience is also displayed on a LED screen or projected onto two adjacent walls. The real-time sound in VR is played through a speaker in the exhibition so both the participants and the immersant can hear.

Figure 8. *LAVIN,* Virtual Reality Art Installation, Siggraph Art Gallery at Los Angeles Convention Center, ©Weidi Zhang.

One participant at a time has the chance to be immersed in VR (see Figure.9), and other participants will possibly be captured by the cameras, which determines the virtual journey of the immersant. Immersants are not given any instructions or visual cues to help them decide what to do, the AI system enables intuitive user interactions which avoid the learning curve caused by using VR controllers.
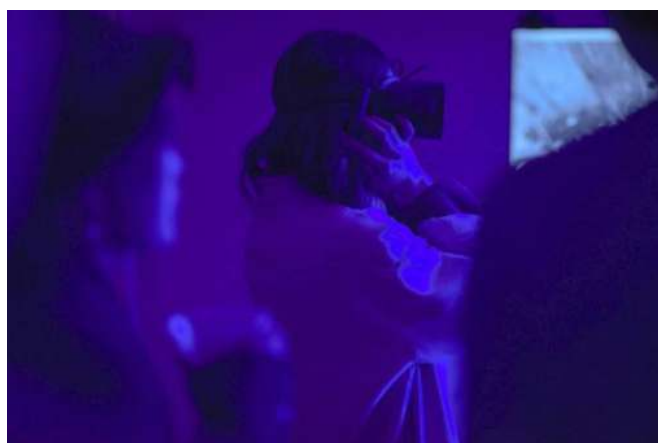

Figure 9. *LAVIN,* An Participant Experiencing the VR, SwissNex Gallery, ©J Astra Brinkmann.

During the exhibition, participants usually immerse themselves in the VR world for around 10 to 15 minutes. Immersants are engaged in the VR experience using a headset and being automatically transported to different locations determined by AI's observation. The

immersants wander within the space, investigate the meaning of the abstract sculptures of objects, moving patterns, and abstract sounds they encounter, and free their imaginations of this artistic interpretation of an AI's mind.

Discussion: Symbiosis, Metaphors, and Cognitive Assemblage

LAVIN emphasizes the symbiosis between humans and machines through contributions from artists, machines, and audiences. This cognitive assemblage formalizes a multi-layered decision-making process where the cultural layer (generated by humans) and the computer layer (generated by the software) mutually influence each other, resulting in a blend of human and machine meanings.

This fluid network begins with critical and cultural questions that determine the information collected for an AI system to learn, which forms the first layer: a cultural layer. The intelligent system makes decisions based on its prior knowledge and observation of images, forming the second layer: a computer layer. The artist's aesthetic decisions and audiences' real-time feedback forms the following cultural layers. Each layer is interconnected and adaptive. For example, participants' real-time feedback is constantly captured by the camera (imaging device), which affects the second layer of the decisions made by the AI system. The interconnection between different layers contributes to the entanglement of culture and computer layers, forming a cognitive assemblage in system design.

N. Katherine Hayles stated in Cognitive Assemblages: "Technical Agency and Human Interactions: In a cognitive assemblage, humans and technical systems are interconnected with each other. The interactions involved human cognition, which includes consciousness, the unconscious, the cognitive nonconscious, and the perceptual system. In addition, the decisions made by humans will, to some extent, determine how the technical system operates." [9]

In *LAVIN*, the output of the technical system will also, many times, affect humans' decisions, which leads to a cognitive assemblage in the art experience, the feedback loops steer human-machine collaboration to a "more life-affirming practices as we move toward a future in which technical agency and autonomy become increasingly intrinsic to human complex systems" [10]. For instance, the participants who are captured by the

camera in the art installation can see the real-time projection of the virtual world. When they see the machine's recognition of them leads the immersant travel to a specific object (for example, a cactus), the participants tend to interact with our system by placing different things in front of the camera and see machine's real-time interpretation changing the route in VR. Therefore, when we design a technical system, we are partially designing ourselves [10].

LAVIN provides an art experience that is not only designed as a cognitive assemblage but also as a spatial assemblage, where metaphors are used to build poetic meanings and connections are employed to build metaphors. We emphasize a metaphorical approach to building connections between visual materials and forms while embracing chance and choice operation. The metaphorical approach to building organizations of media elements (text, images, sounds) consists of automatic combinations and manual compositions. *LAVIN* manually assembles the spatial arrangements of virtual objects our AI system can recognize in 3D virtual space. The contrast, directions, and scale of objects are intentionally exaggerated to illustrate the surreal and bizarre assemblage that suggests the 'AI's brain' within virtual reality. *LAVIN* is not only about the co-creation of experience but also about co-authoring the metaphors and connections, which emphasizes change and instability, fluidity, and exchangeability.

Future Direction

When we exhibit this artwork, the most common question raised by participants is why we selected this list of 50 daily objects. Are there any special meanings associated with these objects? The curiosity raised by participants motivates us to develop *LAVIN* further in the future. We plan to create a more precise and curated collection of objects rather than selecting them randomly. As a result of the careful curation of objects, our trained AI system will develop a worldview that will inspire the interesting and critical questions of human-machine reality.

Conclusion

In this paper, we introduce the background, design methodology, and future direction of an AI-guided VR experience *LAVIN*. In addition, this paper discusses the metaphorical approach to present cognitive assemblage in system design and spatial assemblage in

worldbuilding. *LAVIN* provides an art experience to visually explore one understanding of a neural network in which the real world is mapping to fifty daily objects. The AI system constantly observes and analyzes the surroundings via a camera and outputs one word as a semantic interpretation. This word navigates an audience into a virtual world consisting of abstract fluid structures of these fifty daily objects that *LAVIN* can understand. *LAVIN* tends to evoke audiences' awareness regarding values, vulnerability, and beliefs within the context of AI. It also provides novel research to connect VR and AI through designing an intelligent navigation system for an immersive art experience.

Acknowledgment

References

1 Zaidi Syed Sahil Abbas, Mohammad Samar Ansari, Asra Aslam, Nadia Kanwal, Mamoona Asghar, Brian Lee, "A survey of modern deep learning based object detection models." *Digital Signal Processing* 2022, 103514.

2 Name deleted to maintain the integrity of the review process.

3 Seymour Chatman, "What Novels Can Do That Films Can't" *Critical Inquiry* 7, no. 1, 1980, 121-140. https://doi.org/10.1086/448091.

4 Akten Memo, Rebecca Fiebrink, Mick Grierson, "Learning to see: you are what you see", In *ACM SIGGRAPH 2019 Art Gallery (SIGGRAPH '19),* Association for Computing Machinery, New York, NY, USA, Article 13, 2019, 1–6, DOI: https://doi.org/10.1145/3306211.3320143.

5 Oliver Grau,"Into the belly of the image: historical aspects of virtual reality", *Leonardo* 32, no. 5, 1999, 365-371.

6 Chesak Renel, "CNN-vs-ResNet50-trained-on-20-images", Kaggle, 2020, Accessed Dec 8th, 2022, https://www.kaggle.com/code/rchesak/cnn-vs-resnet50-trained-on-20-images.

7 Brett Koonce, "ResNet 50", In *Convolutional neural networks with swift for tensorflow,* 63-72, Apress, Berkeley, CA, 2021.

8 "Fish Magic," Philadelphia Museum of Art, Accessed June 1st, 2021, https://www.philamuseum.org/collection/object/51027.

9 N. Katherine Hayles, "Cognitive Assemblages: Technical Agency and Human Interaction", *Critical Inquiry* 43, no. 1, 2016, 32-55. https://doi.org/10.1086/688293.

10 N. Katherine Hayles, "The Cognitive Nonconscious: Enlarging the Mind of the Humanities", *Critical Inquiry* 42, no. 4, 2016, 783-808. https://doi.org/10.1086/686950.