SOVIA: Sonification of Visual Interactive Art

Lauryn Gayhardt and **Margareta Ackerman**
Department of Computer Science & Engineering
Santa Clara University
{lgayhardt,mackerman}@scu.edu

Abstract

This paper presents SOVIA, an interactive system that endows Claude Monet's art with responsive auditory experiences. SOVIA uses computer vision trained on Monet's artwork to take the user "into the painting." When the user interacts with a digital version of Monet's landscapes, their mouse positions are mapped to sounds that artistically represent the objects that the user is currently exploring in the art. These interactive musical journeys have the potential to make classical art more captivating for modern audiences.

Introduction

In recent years, entertainment has been becoming progressively more interactive - from social media, art, and education, there is an effort to engage users beyond mere consumption. This opens up the challenge of how to endow classical art forms with new layers of interactivity to engage modern audiences, while retaining the essence of the original art.

Visual art uses color, light, texture, and stroke techniques to convey the mood, tone, and meaning of the artwork. Every layer of information aids in expressing the artist's intent. Adding music and sound to a painting can assist in creating more depth, strengthen existing themes, and convert a consumption-based experience to an interactive one.

In this paper, we propose an interactive method for deepening engagement with visual art. We introduce SOVIA (Sonification of Visual Interactive Art), an interactive system that adds an auditory dimension to still art. Our initial version of SOVIA endows Claude Monet's landscape paintings with soft music and nature sounds, which respond to user's mouse position. The aim to bring the user "into the art," letting them experience a self-directed musical journey into Monet's landscapes (see https://github.com/lgayhardt/SOVIA for a demo).

Monet was a French impressionist artist (1840-1926) and the first to paint outside the studio. He aimed to capture "what is seen rather than what is known" (Seitz 2021). SOVIA's sounds added to his paintings amplify the idea that the viewer is experiencing a slice of life.

Instead of trying to accurately reproduce the scene before him in detail, Monet aimed to record "on the spot the impression that relaxed, momentary vision might receive—what is seen rather than what is known, with all its vitality and movement" (Seitz 2021). To reflect this balance of the literal and metaphorical, SOVIA adds an auditory dimension that intermixes real sounds with musical elements.

Using machine learning enabled object detection, SOVIA recognizes objects in the paintings, which are subsequently mapped to sounds. When the user glides their mouse over a hill they will hear sounds of herding bells through the background music. If the user's mouse wonders over flowers, chimes will play, similar to what one may hear in a garden as a soft wind floats by. This mixture of music with realistic and associated sounds creates an experience that mimics realistic elements in the art, while reflecting the gentle artistic reinterpretation of those objects through sound.

The interactive process offered through SOVIA places most of the effort on the machine agent, while giving the user a simple and enjoyable experience that deepens engagement with Monet's landscapes. We hope that the process proposed here will inspire more researcher into how creative machine agents can be used to enliven classical art forms.

Method

SOVIA is written in Python, using Microsoft Azure's Custom Computer Vision Service for object detection and Pygame to handle tracking the mouse movements and playing the sounds. When a user's mouse enters the coordinates of a detected object it will select an associated sound and increase the volume of the sound to an audible level.

Computer Vision

We manually created the training data for the model. The training data consisted of 369 various sized Monet landscape paintings with the tags: flowers, snow, structure (building), grass, mountain, water, sky, and boat. The training set includes 55 tags for flowers, 53 for snow, 158 for building, 161 for grass, 57 for mountains, 212 for water, 57 for boat, and 318 for sky. The model was trained for five hours.

To evaluate the model's performance, we consider precision, which is the likelihood that a tag predicted by the model is correct; recall, the probability that the model found all the objects in a given image, and mean average precision(mAP). Our model has an 80% probability threshold with 62.2% precision, 71.2% recall and 64.8% mAP.

Proceedings of the 12th International Conference on Computational Creativity (ICCC '21)
ISBN: 978-989-54160-3-5

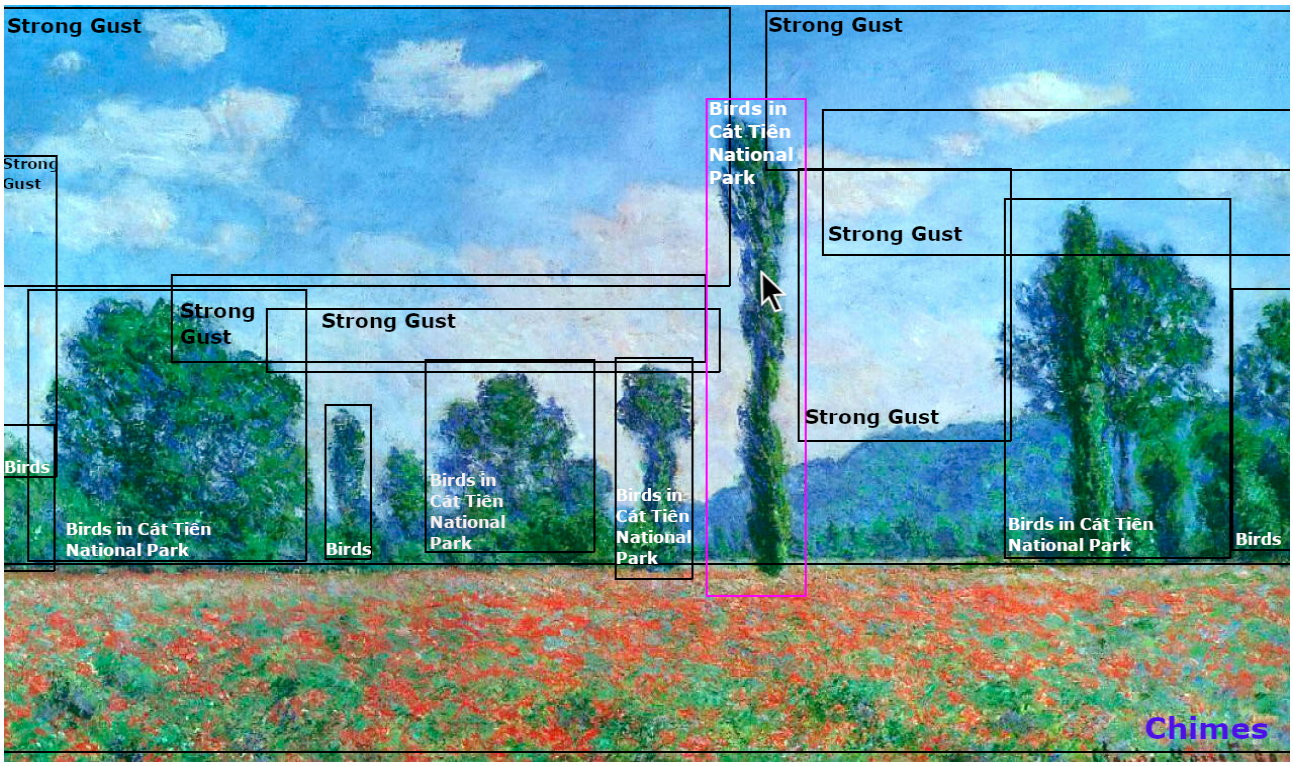

Figure 1: **A visualization of the internal workings of SOVIA**. The bounding boxes detected using computer vision are labeled and mapped to a set of sounds, one of which is played when the cursor enters the box. As the user explores the painting through mouse movements, sounds corresponding to the different bounding boxes are intermixed with the background music. Bounding boxes are not shown to the user. The sky boxes are mapped to sounds of a strong gust of wind, the tree boxes are mapped to bird sounds in Cát Tiên National Park Vietnam, while the flower box is mapped to sounds of chimes. The position of the cursor dictates the active box whose sound is played, visually represented here through highlighting in this image. The painting is of Poppy Field at Giverny (1890) by Claude Monet. Photo Credit: WikiArt

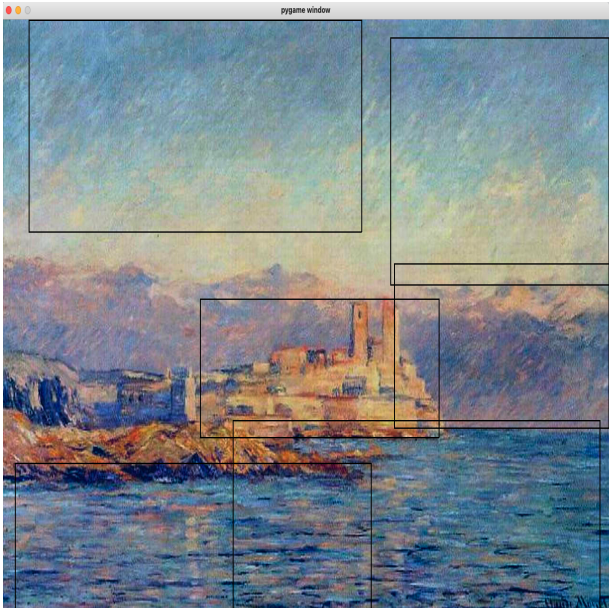

Figure 2: The Castle in Antibes (1888) by Claude Monet with bounding boxes on water, sky, and building objects. Photo Credit: WikiArt

Sounds

The custom vision service uses normalized coordinates with left, top, width, and height to detect objects. In this project normalized coordinates are converted into pixel coordinates to compare the mouse location in Pygame. Each tag has corresponding sounds associated with it, some tags have one sound while others have many. When there is more than one sound associated with a tag the sound to be played randomly selected. All tags are mapped to a Pygame sound channel, which starts at a volume of 0.

When the mouse location intercepts the boundary for a detected tagged box it increases the volume for the sound channel for that tag until it reaches a clearly audible volume which it will remain at until the mouse is moved outside the boundary box. When the mouse is moved outside the box the volume starts to decrease until it is 0 again. Increasing and decreasing the volume allows for sounds to continuously play, so when the mouse intercepts the bounty for a box, the sound does not start over again.

If there is more than one of a certain type of tag detected, the same sound is still played for that tag. For example, if an image has two tags of trees, one of the many tree sounds will be selected to be played for all tree tags. When the mouse intercepts one of the tree tag boxes, the volume for that sound will be increased.

Sounds and music are royally free and were obtained from Zapsplat, Freesound, Mixkit, and Avosound. Please see Figure 3 for a list of object-sound mappings used in SOVIA. Some tag sounds were chosen for their literal representation of the tagged feature while others are more abstract. For example, Chimes were used for the flower tag to create the feeling that one is walking into a garden and hears chimes

Tag	#	Sounds
Trees	10	Forest sounds with birds
Sky	5	wind sounds
Flower	1	Chimes
Grass	1	Cicadas with distant birds
Water	7	Sea,stream and underwater sounds
Mountain	2	Goats and herding bells
Snow	3	Footsteps in the snow
Boat	4	Boat swaying in water, creaking wood and paddling
Building	1	People chatting in Polish

Figure 3: SOVIA's object-sound mappings. The left column represents objects found in Monet's painting, the second column corresponds to the number of sounds that SOVIA offers for that object, and the final column describes the nature of the sounds given to the object.

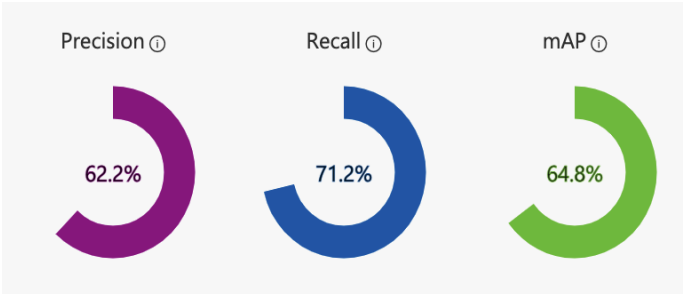

Figure 4: Our computer vision model with 80% probability threshold performance metrics. Photo Credit: Microsoft's Custom Vision Portal

blowing in the wind. Chatting in Polish is used in the building to give the sense of a lively conversation taking place within the house. Goats are heard for the mountain tag because it is one of the many animal sound that one may hear on a mountain top.

Related Work

Related work has considered the conversion of art to music and vice versa, often focusing on creating novel works rather than enhancing existing art works. Other related work offered co-creative experiences which allow the user to concurrently create new music and art. We share several examples here.

MetaSynth (U & I Software 2021) lets users create music from images . It allows for pixels to be drawn or imported onto a digital canvas, and uses the RGB color data from pixels to affect and produce sound. It generates audio using the brightness of the pixels to control to amplitude, pitch bases on the pixel's Y-axis position, and the red and green color components to pan the audio to the left or right speaker channel (Pitman 2009). Metasynth differs from SOVIA in that it generates audio at the pixel level of an image vs detecting objects within an image that map to different sounds to be

played.

PhotoSounder (Rouzic 2020) is a graphical and audio editing software and synthesizer that allows the user to make music from images. It is similar to MetaSynth, but gives the user greater control over sound mappings and runs on both Windows and Mac.

Pixelsynth (Jack 2016) is a browser based synthesizer inspired by the analog ANS Synthesizer that creates sounds from images and drawings. It uses the grayscale version of an image and turns it into a sine wave by having the white in the picture represent a note that is on, transparency representing velocity of the note and location for pitch(Arblaster 2016).

A system by Joana Teixeira and H. Sofia Pinto (Teixeira and Pinto 2017) takes an image as an input and generates music by relating visual features to musical ones. Conversely, a system by Luís Aleix, H. Sofia Pinto, and Nuno Correia (Aleixo, Pinto, and Correia 2021) taking in music which it uses to generate an abstract image.

SOVIA differs from the systems above because it is not transforming one art form into another, or offering a co-creative experience that creates novel art and music, but rather amplifying a given artwork by adding a musical interactive experience.

Related Digital Art

Digital artists have also experimented with interactive integration of sounds and visuals through online installations. For example, Joe Hamiliton's Indirect Flight is an interactive web art that displays a layered collage of landscape images set to a realistic soundscape of wind aeroplanes and other urban noises by J.G Biberkopf. "As you pan across the terrain like Google Maps the layers move at different speeds giving the illusion of depth, constantly changing what is hidden and exposed." (Hamilton 2015).

Rafaël Rozendaal's Sunrise/Sunset allows users to interact with circles that represent the rising and setting of the sun in New York City on the Whitney Museum of American Art website creating a piece of abstract art. "As visitors to whitney.org move the cursor over the black or white circle obscuring the web site, they cast spinning shadows or light over the page that obscure or expose its content" (Whitney Museum 2017).

Conclusions and Future Work

Creative machine agents can add interaction to traditionally consumption-only visual art experiences. The interaction can result in a more immersive experience and perhaps invite viewers to more deeply engage in meaning formation.

SOVIA adds an auditory component to traditional visual art, particular the works of Monet. Adding acoustics that blend realistic and symbolically-related sounds with soft music pairs well with Monet's style - inviting viewers to a self-directing sonified experience of Monet's landscapes.

Future work will carry out user studies comparing the consumption of artwork with and without the interactive musical dimension added by SOVIA. We would also like to expand SOVIA to other artists, as well as allow users to upload their own images which the system will endow with an interactive musical dimension.

References

Aleixo, L.; Pinto, H. S.; and Correia, N. 2021. From music to image a computational creativity approach. In *Artificial Intelligence in Music, Sound, Art and Design: 10th International Conference, EvoMUSART 2021, Held as Part of EvoStar 2021, Virtual Event, April 7–9, 2021, Proceedings 10*, 379–395. Springer International Publishing.

Arblaster, S. 2016. The noise of art: Pixelsynth can turn your images into music for free. https://www.musicradar.com/news/tech/the-noise-of-art-pixelsynth-can-turn-your-images-into-music-for-free-638423.

Hamilton, J. 2015. Joe hamilton – indirect flights. https://indirect.flights/.

Jack, O. 2016. Pixelsynth. https://ojack.xyz/articles/pixelsynth/index.html.

Pitman, T. 2009. An introduction to metasynth. https://music.tutsplus.com/tutorials/an-introduction-to-metasynth–audio-2425.

Rouzic, M. 2020. Photosounder: Image-sound editor amp; synthesizer. https://photosounder.com/.

Seitz, W. C. 2021. Claude monet. *Encyclopedia Britannica*.

Teixeira, J., and Pinto, H. S. 2017. Cross-domain analogy: From image to music. In *Proceedings of the 5th International Workshop on Musical Metacreation*.

U & I Software. 2021. Metasynth + xx. https://uisoftware.com/MetaSynth/.

Whitney Museum. 2017. Rafaël rozendaal: Sunrise/sunset. https://whitney.org/exhibitions/rafael-rozendaal.