

AIMC 2024 (09/09 - 11/09)

Deep Drawing: An Intermedia AI Co- Performer

Julie Zhu, Erfun Ackley, Zhiyu Zhang, John Granzow

Published on: Mar 23, 2024

URL: <https://aimc2024.pubpub.org/pub/859kww47>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Title

Deep Drawing: An Intermedia AI Co-Performer

PubPub Link

<https://aimc2024.pubpub.org/pub/859kww47/draft?access=y127wioa>

Project Description

Deep Drawing is an intermedia performance about the sound of drawing. Using the live audio input captured from the human performer's drawing/writing, the AI performer generates in real-time a prediction of what the drawing looks like. The 10 minute performance consists of a human performer making aurally interesting drawings on a wooden surface outfitted with four contact mics, exploring timbre and rhythm and repetition (Figure 1). The AI performer translates the nuances and variations in the audio into corresponding visual movements and marks on a projected digital canvas. During the performance, the real drawing and the predictive drawing are mixed in OBS to highlight the successes and failures of the machine.

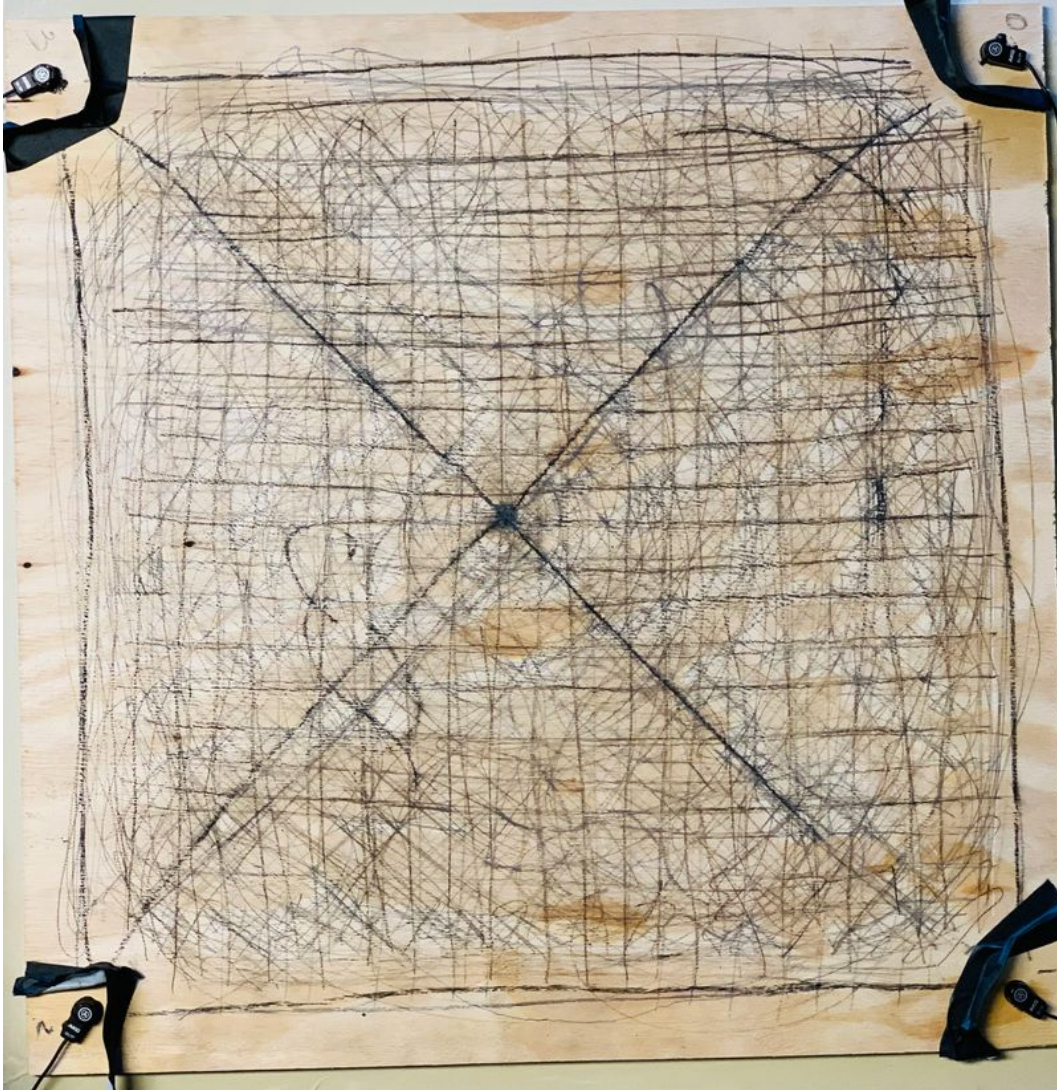


Figure 1: Wooden surface (2x2 ft) setup with four contact mics.

Our system leverages convolutional neural networks (CNNs) to generate real-time pen positions from audio input, enabling a novel drawing application. The model processes four-channel audio spectrograms through a series of convolutional and fully connected layers. The CNN outputs continuous 2D coordinates, translating complex audio patterns into fluid pen movements. Trained on a dataset of synchronized audiovisual pairs, the model learns to map subtle audio cues to precise spatial locations. These predicted coordinates are subsequently streamed to a web-based JavaScript application, where they are rendered in real-time, forming the foundation for a visual overlay that complements the live audio performance (Figure 4). This design unveils the mutable aesthetics of deep learning, bridging the gap between auditory perception and visual expression (Figure 2 and 3).

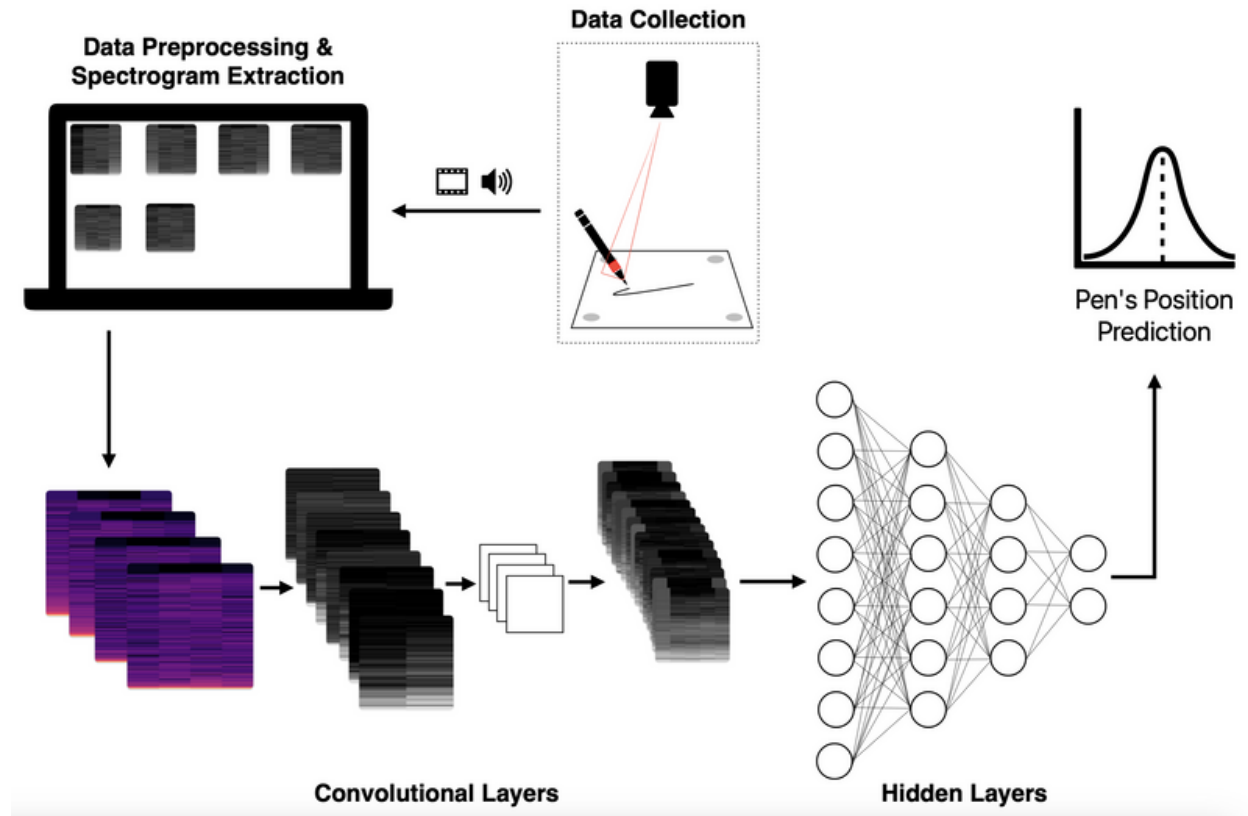


Figure 2: CNN Framework for Real-Time Audio-to-Pen-Position Mapping: This flowchart outlines a model that commences with the synchronous recording of audio and video of drawing on a wooden surface, proceeds through data pre-processing, and culminates in the extraction of spectrograms. A CNN takes the spectrograms from each of the four contact microphones to localize the sound, where the output is a continuous set of real-time (x, y) coordinates, which are then visualized through a JavaScript front-end application, effectively transforming auditory stimuli into corresponding visual movements.

We are interested in the intersection of sound and visual art through a very specific lens and often ignored part of our daily life: the sound of writing and drawing. With deep learning, the intricate and soft noises of graphite on paper can come to life through the machine's unexpected and possibly absurd visual interpretations of the sound. By proposing such an impractical use of Music AI, we hope to explore an alternative philosophy of human and machine creativity.

Type of submission

As our performance centers around video and would operate best with immersive audio, Performance 2, at an Oxford University performance space with a flexible stage and electronics/projected visuals set-up, seems the most suitable. \textit{Deep Drawing} will provide its own human performer.

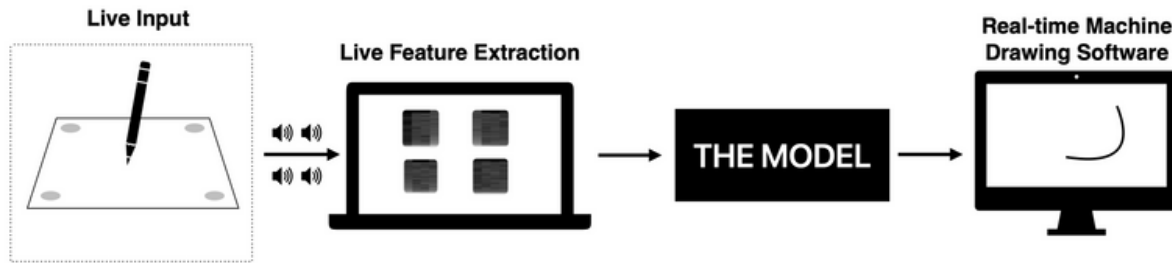


Figure 3: The Live Setup

Technical/Stage Requirements

- 4 contact mics attached to a 2 foot square wooden surface are routed through an audio interface into a computer that outputs 4 spatialized audio channels through the same interface.
- The wooden surface sits on a table where a mobile phone captures the live drawing that is also fed into the computer which live-mixes the generated video and the live video.
- Four contact mics map to four speakers to preserve the sense of an audience being surrounded by the sound of drawn circles. We hope the venue can provide speakers, cabling, an audio interface with 4 pre-amps, the requisite outs, a table and a projector.
- The performer will bring a computer, a document camera, 4 contact mics, a wooden box, drawing utensils, and an audio interface if necessary.



Figure 4: Sample video output with AI visual co-performer overlaid on live video.

Program Notes

Imagine your signature, its particular loops with its particular rhythms. Imagine writing your signature over and over until it becomes a loop and a beat. Hold onto that sound. Now imagine a machine that tries to guess your signature just from the sound. With enough mics and training, would it, could it, reproduce your signature?

Deep Drawing is an intermedia performance that explores the sound of drawing and writing through a deep learning model. An AI co-performer interprets the musical gestures of the human co-performer drawing on an amplified wooden surface to create projected visual drawings corresponding to the sound. Trained on the idiosyncratic drawings of performer and composer Julie Zhu, a research team at the University of Michigan (Julie Zhu, Erfun Ackley, Zhiyu Zhang, John Granzow) used neural networks to design a system to test the machine's capabilities of bringing the intricate and soft noises of graphite on wood to visual life.

Media

Video score: <https://youtu.be/JFIQEoiz8SI>

Ethics Statement

All data used in this research was generated by the researchers. Our code uses open-source libraries in Python and JavaScript. We tried to reduce our environmental impact by minimizing computing resources.

Acknowledgements

- This work is supported by the Performing Arts Technology Department and by the ADVANCE program at the University of Michigan.