Synthaesthetic Art: Human-Machine Creative Collaboration

Justin Baird
Tesseract.art
justin@tesseract.art

Ivy Chen
Tesseract.art
ivy@tesseract.art

Jookyung Song
Seoul National University
chsjk9005@gmail.com

MooKyoung Kang
Xorbis
nostone@gmail.com

Richard Savery
University of Canberra
saveryrichard0@gmail.com

Abstract

Synthaesthetic Art is a human-machine collaborative framework combining live perception, generative AI, and robotic execution. Presented as a live installation at the **Super AI Conference 2025**, the system captured portraits of attendees, transformed them into stylised caricatures using a diffusion model that was fine-tuned on artist-in-residence Ivy Chen's works, and rendered them on canvas with acrylic paint via a robotic arm. Unlike conventional AI-driven art systems, our Synthaesthetic Art system was designed to preserve and amplify artist authorship Artist authorship is preserved through stylistic training, selective curation, and real-time control. This work introduces the conceptual foundations of Synthaesthetic Art and explores how hybrid creativity can expand agency, re-frame authorship, and deepen emotional expression through machine augmentation.

1 Introduction

What happens when a robot doesn't just assist the artist, but becomes part of the artwork?

Synthaesthetic Art is a contemporary movement that explores how generative algorithms, sensory inputs, and robotic systems can be integrated into live, co-creative processes. Rather than replacing the artist, Synthaesthetic Art positions AI and robotics as collaborators, amplifying human intent while challenging traditional notions of authorship and aesthetic form. This paper is grounded in a collaborative project first demonstrated at the **Super AI Conference 2025 in Singapore**, where portraits of attendees were captured, stylised via a diffusion model fine-tuned on artist-in-residence Ivy Chen's caricatures, and painted by a robotic arm. The process introduced Ivy's distinctive style and showcased a new form of hybrid creativity in which human authorship and machine execution are intertwined.

The movement is grounded in six core principles: **Machine-Augmented Creativity**, which emphasises the machine's role as an extension of human expression; **Hybrid Intelligence**, wherein decisions are shaped by both human intuition and computational inference; **Temporal and Spatial Fluidity**, which embraces adaptive, real-time creation; **Sensory Fusion**, blending modalities like sound, vision, and gesture into unified compositions; **Algorithmic Aesthetics**, exploring emergent patterns and computational beauty; and **Decentralised Authorship**, which questions the notion of a singular artistic origin.

This paper presents a system that embodies these principles, transforming facial images into acrylic portraits while preserving and extending the artist's authorship. We situate this work as both a

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.

technical contribution and an artistic provocation, demonstrating how AI systems can be embedded in creative workflows in ways that remain expressive, ethical, and deeply human.

2 Early Explorations in Sound and Vision

Our initial studies connected live music with robotic painting through a real-time audio-robot pipeline. Max/MSP analysed features such as pitch and onset density, triggering brush gestures via OSC messages to a Python-controlled robotic arm. Figure 1 illustrates examples from these formative experiments, showing different ways we connected music and robotic drawing. This early setup raised the central question of our practice: where does authorship lie when sound becomes image? (Savery et al., 2022)

A key turning point came in a performance with vocalist *Nyali*, where we introduced a vision layer. A camera tracked her facial position and micro-expressions, while audio features continued to shape brushstroke timing and type. OSC sliders modulated speed, jitter, and stroke behavior. The result was a responsive, multimodal system that performers described as "the painting singing back."

This collaboration revealed two core lessons: (1) mapping embodied cues such as gaze enables clearer performer agency; (2) maintaining human supervision keeps the robot an expressive instrument, not an autonomous painter. These insights shaped our current Synthaesthetic pipeline. (Savery and Baird, 2024)

Earlier prototypes included violin-controlled sketches, a large-canvas system with automated brush swapping, and group performances where musicians signed the resulting artwork—early attempts that explored the boundaries of shared creative authorship.

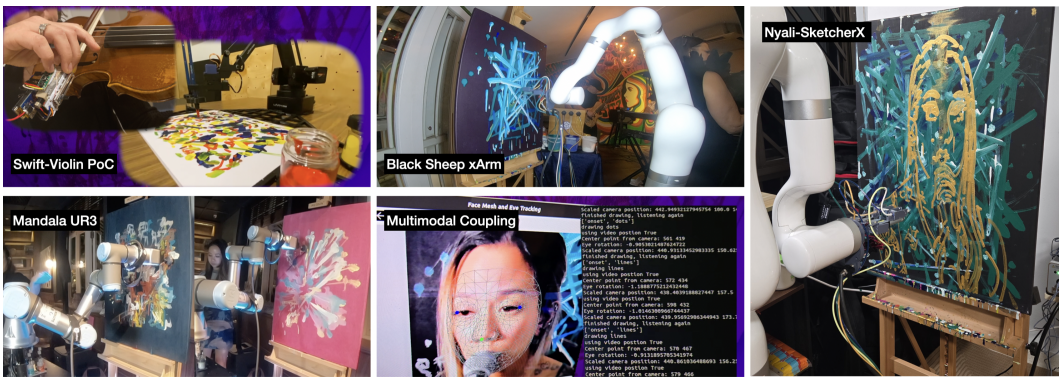

Figure 1: Examples of early experiments linking music and robotic drawing. Left and middle: music-driven robotic sketches. Right: performance with vocalist *Nyali*, where her singing generated the painted background and her live portrait was drawn at the end of the performance.

3 System Design and Technical Pipeline

3.1 Image Preprocessing and Prompt Generation

To produce stylistically consistent portraits, we developed a streamlined capture and preprocessing pipeline aligned with our artist-adapted diffusion model. Portraits were taken using a fixed DSLR setup, then automatically aligned and cropped to match the forward-facing composition typical of Ivy Chen's caricatures.

We applied Meta's Segment Anything Model v2 (SAM2) (Ravi et al., 2024) to remove background and clothing details, reducing stroke complexity and emphasising facial features for improved stylistic fidelity and faster robotic execution.

Because our fine-tuned model builds on Stable Diffusion, high-quality text prompt generation is essential. To automate this, we used Microsoft's Florence-2 (Xiao et al., 2024) to extract visual features—such as age, gender, hairstyle, and accessories—which were rephrased by a LLaMA-based

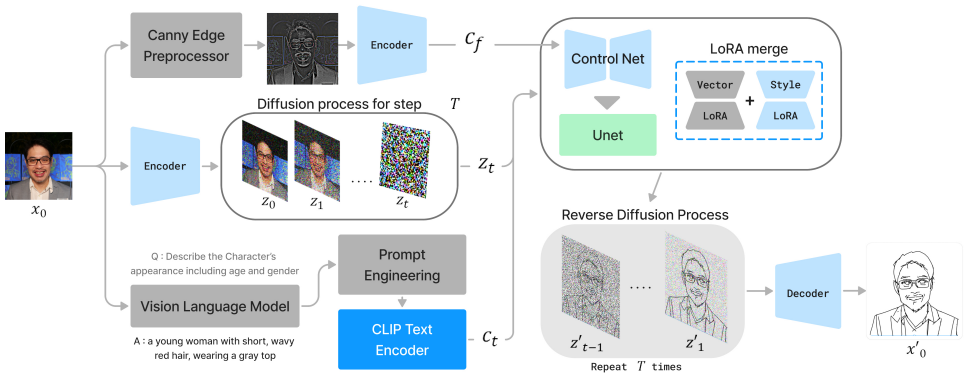

Figure 2: Technical pipeline showing portrait capture, image preprocessing feature extraction with canny edges, and diffusion-based stylisation.

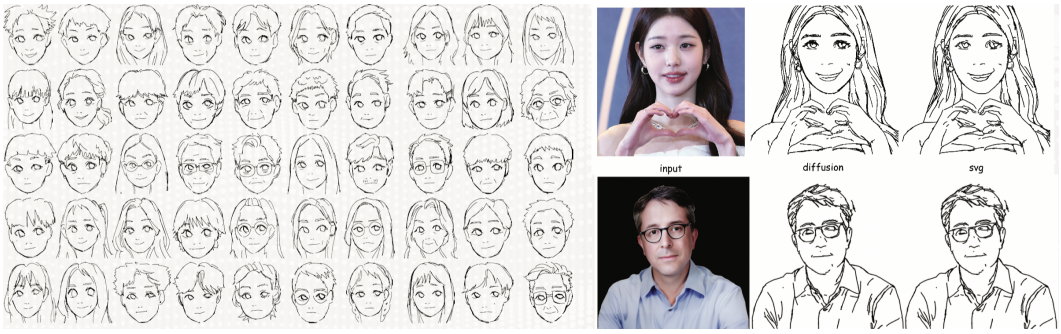

Figure 3: Left: hand-drawn caricatures by Ivy Chen used to train the model. Right: model inference pipeline – input photograph of a person (left), diffusion-generated stylised image (middle), and final vectorised SVG output (right) prepared for robotic execution.

language model into descriptive prompts. This combination enabled high-quality, personalised generation without manual intervention.

3.2 Diffusion Model and LoRA Training

To adapt Stable Diffusion to Ivy Chen's caricature style, we fine-tuned it using two Low-Rank Adaptation (LoRA) modules: a *Style LoRA* trained on 50 of Ivy's hand-drawn portraits (examples shown in Figure 3, left), and a *Vector LoRA* introduced in Song et al. (Song et al., 2024). It is optimized for clean, continuous strokes suited to robotic execution. This dual-LoRA approach preserved stylistic fidelity while ensuring vectorization compatibility.

The dataset included diverse demographics to support generalisation. Style images were paired with regularisation images to maintain structural coherence. To enhance alignment, we incorporated context consistency loss and ControlNet with canny-edge conditioning (Zhang et al., 2023).

Prompt embeddings used CLIP (Radford et al., 2021) to align textual inputs with generated features. The resulting model reliably produced stylised, vector-friendly caricatures, enabling efficient SVG conversion and robotic rendering via continuous Bézier strokes—crucial for fluid, human-like execution.

3.3 Vectorisation and Robotic Rendering

The selected diffusion output is converted to scalable vector paths (SVG). We approximate primitives (paths, lines, curves, polygons) as continuous Bézier segments and normalise them to the physical canvas.

Each path is then executed stroke-by-stroke by an xArm robot. We approach, draw, and retreat with controlled lift motions to avoid smearing, and expose real-time parameters over OSC from Max/MSP. These modifiers let sound or performance cues adjust stroke density and continuity without breaking geometric fidelity.

Brush changes, pressure offsets, and per-stroke safety margins are handled in software, so the arm behaves like a responsive prosthetic. The artist can pause, skip, or replay individual strokes, keeping curatorial control over every mark. In practice, this makes the robot an expressive executor, not an autonomous painter.

4 Artistic Agency and Direction

The creative foundation of the Synthaesthetic system is grounded in the work of Ivy Chen, the artist-in-residence. Her hand-drawn caricatures formed the primary training corpus for the diffusion model, and her aesthetic decisions shaped both the stylistic parameters of the generation process and the behavior of the robotic rendering system.

To construct the dataset, Chen produced over fifty black-and-white line drawings, each reflecting her distinct visual language. These caricatures encompassed a broad range of demographic features—including variations in facial structure, hairstyle, age, and accessories—to promote stylistic generalisation while maintaining authorial coherence. Technical constraints for robotic painting informed the drawings: shading was omitted, filled regions avoided, and features such as eyes were rendered with open lines or hatching. The dataset evolved through iterative cycles of testing with close attention to line clarity, demographic balance, and stroke feasibility.

This preparation reflects the principle of **Algorithmic Aesthetics**, wherein visual form is shaped through computational constraint and artistic abstraction, and **Machine-Augmented Creativity**, as the artist's own style becomes both the training signal and the creative boundary within which the machine operates.

Following model fine-tuning, Chen retained curatorial authority over all system outputs. She conducted stylistic evaluations of generated portraits, selecting those that aligned with her visual intent and rejecting or refining those that did not. During robotic execution, she exercised real-time control to modulate stroke attributes. In this manner, authorship was not merely pre-configured but remained dynamic and embodied during the act of rendering.

This interactive control loop exemplifies **Hybrid Intelligence**, whereby the system's generative and mechanical components are continuously shaped by human intuition and aesthetic judgment. It also enacts **Temporal and Spatial Fluidity**, enabling the artwork to evolve over time and across layers of abstraction—from image to vector to brushstroke.

Moreover, the multimodal integration of visual, physical, and performance-based inputs reflects the principle of **Sensory Fusion**, allowing audio features, gesture, and live input to co-shape the final composition.

Chen's role extended across the entire creative pipeline: dataset curation, style definition, generative evaluation, and real-time robotic orchestration. This expansive authorship challenges conventional notions of singular authorship in computational art and directly aligns with the principle of **Decentralised Authorship**, wherein the final artefact emerges from a distributed network of human, algorithmic, and mechanical contributors, with the artist acting as a conductor of this hybrid ensemble.

This framework defines the function of the **Synthaesthetic Artist**: one who authors the work, but the generative conditions, the operational logic, and the expressive dynamics of the system itself. Chen's influence is embedded not only in line, form, and style, but in the computational grammar and mechanical gestures that bring each portrait into being. Her authorship is both visible and systemic; a signature of intent encoded across every layer of the creative process.

5 Foundations of Human-Machine Collaboration

The development of our real-time human-machine collaborative art system is fundamentally shaped by the core principles of the Synthaesthetic Art movement. These principles—Machine-Augmented Creativity, Hybrid Intelligence, Temporal and Spatial Fluidity, Sensory Fusion, Algorithmic Aes-

Figure 4: Ivy Chen holding a completed portrait painted, while in the same scene the robot continues to generate a new portrait at SuperAI conference in Singapore.

thetics, and Decentralised Authorship—provide both philosophical and technical scaffolding for our design. Below, we articulate how each principle is practically embodied within our system and workflow.

5.1 Machine-Augmented Creativity

Our system positions artificial intelligence not as a replacement for human creativity, but as a means of augmenting it. Ivy Chen, the artist-in-residence, retains creative control throughout the process—from dataset curation and style definition to output validation and robot supervision. The generative diffusion model, trained on her caricatures, functions as an extension of her visual language. The robot acts not as an autonomous painter but as a prosthetic executor of her artistic decisions. In this way, the machine serves to expand Ivy's expressive reach, enabling high-speed, stylistically consistent production while remaining subordinate to her authorship.

5.2 Hybrid Intelligence

The system exemplifies a dynamic interplay between human intuition and machine computation. The generative pipeline requires both algorithmic inference (e.g. vision-language embeddings, prompt synthesis, vectorisation) and real-time human judgment. For example, while Florence-2 and LLaMA construct initial prompts from visual features, the final output is shaped through human-led curation, allowing Ivy to reject or refine undesirable results. Real-time robotic rendering is further modulated via OSC controls exposed in Max/MSP, enabling intuitive performance-like interaction during execution. This co-creative structure embodies the hybrid intelligence model at the heart of Synthaesthetic Art.

5.3 Temporal and Spatial Fluidity

Unlike fixed, linear creative pipelines, our system embraces fluidity in both time and space. Paintings evolve through iterative processes—feedback loops between generation, human evaluation, and robotic execution—rather than a one-directional flow. The use of real-time inputs such as audio features or facial tracking data allows the artwork to respond temporally to live stimuli. Spatial fluidity is achieved through vector-based rendering and parameterised robotic motion, enabling adaptive composition adjustments without rigid constraints. These qualities align with the Synthaesthetic ideal of living, adaptable artworks that exist across multiple temporal and representational states.

5.4 Sensory Fusion

Our early explorations integrated sound (pitch, onset density) and visual cues (facial expression, gaze) into a single synthaesthetic canvas, translating auditory and visual stimuli into gestural strokes. The

ongoing use of OSC signals to perturb robotic behaviors—altering speed, stippling, or pressure based on sound or movement—blurs traditional boundaries between modalities. This multimodal mapping reflects the movement's aim to blend senses in ways inspired by neurological synesthesia, where inputs in one modality produce responses in another. Our system literalises this fusion, turning song into line, emotion into form, and presence into composition.

5.5 Algorithmic Aesthetics

The artwork generated by the system embraces complexity, imperfection, and the emergent properties of machine learning. The LoRA-trained diffusion model captures not just Ivy's stylistic features but also introduces subtle variations—emergent quirks, abstraction, and character—that challenge normative standards of beauty. Rather than seeking photorealism or technical polish, we explore an algorithmically-mediated visual language that values stylisation, abstraction, and the uncanny. These properties resonate with the algorithmic aesthetic sensibility of the Synthaesthetic movement, where meaning and form are discovered through generative processes and computational play.

5.6 Decentralised Authorship

The system challenges the traditional notion of singular authorship. While Ivy's role is primary, the system's outcomes are shaped by contributions from model developers, performers (e.g. Nyali), technical collaborators, and even the viewers whose presence influences real-time execution. The curated dataset, open-source components (e.g. SAM2, ControlNet), and community-driven design also reflect an ethos of shared creative ownership. This decentralisation of authorship is a key tenet of Synthaesthetic Art, which embraces the plurality of creators, both human and non-human, in shaping the final work.

6 Conclusion and Future Directions

Synthaesthetic Art is not a single system or technique, but a broader conceptual movement, one that reimagines creativity as a distributed, co-authored process between humans and intelligent machines. It challenges the traditional hierarchies of artistic production by blending sensory modalities, decentralising authorship, and treating generative algorithms and robotic systems as active participants in the creative act.

The system we present in this paper is a concrete realisation of these ideals. It demonstrates how an artist's intent can be embedded into a generative model, how human oversight can guide machine execution, and how live performance data can be translated into physical expression through code and motion. Yet this system is only one instantiation, only one articulation of a much larger set of questions posed by Synthaesthetic Art: What does it mean to co-create with a machine? How do we preserve artistic voice in computational pipelines? And how can technology be harnessed to deepen, rather than dilute, human expression?

Looking ahead, we aim to explore more diverse forms of real-time interaction including gesture, gaze, and audience input to broaden the system's adaptability to multiple artists, styles, and modalities. As machine capabilities evolve, so too must our frameworks for collaboration, ethics, and creative agency. Synthaesthetic Art offers one such evolving framework: not to replace the artist, but to extend their reach and re-imagine what art can become when human and machine think, see, and feel together.

Acknowledgements

We thank Enya "Nyali" Lim, Eivind Lodemel and Tristan Seow for their live music performance. We thank the Super AI Singapore team for supporting the system demonstration.

References

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In International conference on machine learning, pages 8748–8763. PmLR, 2021.

Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, et al. Sam 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714, 2024.

Richard Savery and Justin Baird. Redefining artistic boundaries: A real-time interactive painting robot for musicians. In 38th Conference on Neural Information Processing Systems (NeurIPS 2024) Creative AI Track, 2024.

Richard Savery, Anna Savery, and Justin Baird. Robotic arm generative painting through real-time analysis of music performance. In Proceedings of the 10th International Conference on Human-Agent Interaction, pages 253–255, 2022.

Jookyung Song, Mookyoung Kang, and Noju Kwak. Sketcherx: Ai-driven interactive robotic drawing220 with diffusion model and vectorization techniques. In 38th Conference on Neural Information Processing Systems (NeurIPS 2024) Creative AI Track, 2024.

Bin Xiao, Haiping Wu, Weijian Xu, Xiyang Dai, Houdong Hu, Yumao Lu, Michael Zeng, Ce Liu, and Lu Yuan. Florence-2: Advancing a unified representation for a variety of vision tasks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 4818–4829, 2024.

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In Proceedings of the IEEE/CVF international conference on computer vision, pages 3836–3847, 2023.