Me vs. You: Wrestling with AI's Limits Through Queer Experimental Filmmaking

Adam Cole
Creative Computing Institute
University of the Arts London
London, United Kingdom
a.cole@arts.ac.uk

Gregor Petrikovič
Independent
London, United Kingdom
gpetrikovic@gmail.com

Mick Grierson
Creative Computing Institute
University of the Arts London
London, United Kingdom
m.grierson@arts.ac.uk

Figure 1: *Me vs. You* **still image. In this AI interpretation of a wrestling match, two bodies merge into an unstable form, blurring the boundaries between conflict and connection.**

Abstract

Me vs. You is a multi-channel video installation that explores the complexities of queer intimacy by co-opting an AI machine vision system. In this work, footage from a wrestling match is transformed through a generative video pipeline into a fluid interaction that oscillates between aggression and tenderness. The initial wrestling footage is fed through an AI depth map network—designed to separate bodies—before being reconstructed using a diffusion video process. Rather than rendering discrete fighters, the system produces unstable, shifting forms that sensually collide and merge, destabilizing a clear reading of the interaction. The work exploits the machine vision system's inability to delineate entangled bodies, challenging computational frameworks of classification and control; instead, it repurposes AI as a tool for poetic ambiguity. Situating *Me vs. You* within experimental filmmaking and AI surveillance debates, this paper examines how emerging technologies can disrupt narrow modes of machine perception and proposes more expansive ways of seeing.

CCS Concepts

• **Applied computing** → **Media arts**; • **Computing methodologies** → *Image representations*; *Image manipulation.*

Keywords

AI video, experimental filmmaking, machine vision, installation, queer theory

1 Motivation

Advancements in visual technologies have continually shaped human perception and cognition, influencing how we interpret and engage with the world [4, 9, 16]. AI-driven machine vision, now pervasive in surveillance systems, tends to impose rigid visual classifications that flatten human complexity [14]. These systems prioritize segmentation, classification, and quantification, reinforcing deterministic ways of seeing.

Experimental filmmaking has long disrupted fixed visual paradigms, proposing more fluid and ambiguous modes of perception [22]. *Me vs. You* extends this tradition by repurposing a video-translation pipeline to subvert AI's mechanistic approach to vision, transforming its constraints into a poetic exploration of human conflict and connection. By embracing AI's perceptual failures as a generative tool, the work challenges the rigid taxonomies imposed by machine vision, proposing a more expansive, destabilized visual framework. In doing so, it contributes to a broader discourse on *Creativity for Change* [1] and offers new possibilities of queer representation.

Figure 2: Two-channel installation of *Me vs. You.* Video installation preview here: https://youtu.be/_UWX_YqLC84.

2 Artwork Description

2.1 The Videos

In this two-channel video installation, the artists co-opt a generative video pipeline to interrogate the limits of surveillance-oriented machine vision in interpreting human interaction. Central to the artwork is a live-recorded wrestling match between the artists. This footage is processed through the Depth Anything [25] AI depth map network—a machine vision system designed to estimate spatial relationships by assigning each pixel a grayscale depth value. This technology is typically used to segment subjects; however, the close contact between the fighters pushes the model to its interpretive limits causing it to fail in effectively delineating the wrestlers' bodies.

The artists leverage this failure: when the depth maps are fed into a generative video pipeline that attempts to reconstruct the depth maps spatially, the boundaries of the two entities become blurred. Throughout the generated film, the bodies attack, collide, merge, and separate. Instead of a straightforward wrestling match, the video resists a clear interpretation, shifting fluidly between aggression and tenderness, conflict and intimacy.

The installation features two screens: the lower screen reveals the machine's perspective—raw AI-generated depth maps—while the upper screen presents the generative video system's ambiguous reconstruction. The two screens, positioned perpendicularly, juxtapose two distinct modes of vision—one revealing the machine's narrow perception, the other exposing its interpretive instability.

The installation videos are linked in Appendix B.

2.2 The Prints

Complementing the video installation, are two large-scale prints of the pre-processed depth maps. These grids of images are the video frames processed by the Depth Anything AI network. Here, these
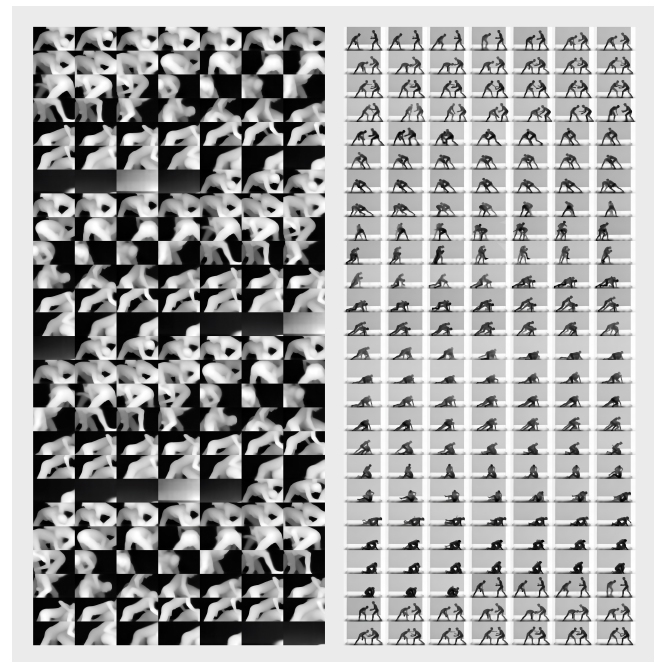

Figure 3: Two accompanying large scale prints (100cm x 200cm) hung back-to-back or across from each other in relation to the video installation.

Me vs. You: Wrestling with AI's Limits Through Queer Experimental Filmmaking

C&C '25, June 23–25, 2025, Virtual, United Kingdom

Figure 4: Example of large-scale hanging print placed opposite the video installation.

Figure 5: *Me vs. You* **at the Tate Britain positioned in dialogue with Francis Bacon's own ambiguous portrayal of male 'wrestlers' [3]. Interestingly, Bacon's work was itself influenced by Eadward Muybridge's** *Some Phases in a Wrestling Match* **[20].**

images are explored as supporting studies to the video work. In many other contexts, these images are considered digital waste: an intermediary step in a larger generative pipeline to be processed by machines as raw data but never seen by humans. Taken out of this mechanized process and visualized in a large-scale grid, they take on their own aesthetic value. Their clinical black-and-white appearance contrasts with the fighters' organic forms evoking an ambivalence that mirrors the themes of the videos. The ambiguous collision, open to interpretation as either aggressive or sensual, underscores the work's central metaphor for the complexity of connection.

2.3 Key Themes

Overall, *Me vs. You* presents a nuanced perspective on queer intimacy, where the boundaries of identity, conflict, and affection collide and blur. The indeterminacy of meaning in the work mirrors queer experience itself—a fluid negotiation of visibility and invisibility, danger and safety, self-loathing and acceptance [7, 17]. Queer identity and relationships resist fixed classification, reflected in how the artwork unsettles the machine's attempt to categorize and fragment bodies within the frame.

By adopting a novel generative process, the work critiques the cold rationality of AI-driven systems and their inability to grasp the complexities of human intimacy, desire, and conflict. Paradoxically, the AI's failure to accurately interpret the wrestlers' entanglement produces a more evocative aesthetic, capturing the invisible friction and grace at the heart of human relationships. By subverting machine vision's rigid logic, the work transforms AI's limitations into a generative space for creative exploration, reinforcing its potential to challenge and expand dominant modes of perception.

2.4 Exhibition

Me vs. You has been exhibited at the Tate Britain in London [24], SIGGRAPH Asia in Tokyo [8], K48 Gallery in Vienna [13], and Studio Teatrgaleria in Warsaw [10].

3 Process

The videos in *Me vs. You* were created using an innovative video-to-video generative translation process, leveraging tools including Stable Diffusion [23], AnimateDiff [12], ControlNet [27] and IPAdapters [26]. The technical pipeline focused on two main components: composition and content generation.

3.1 Composition: Depth Maps and ControlNet

ControlNet is a generative AI model that uses pre-processed structural information, such as depth maps or pose estimations, to guide video frame generation. This process ensures spatial consistency, motion continuity, and compositional stability throughout the sequence. The pipeline for this project was as follows:

(1) The wrestling footage was recorded using a DSLR camera on a tripod with the fighters in front of a flat screen.
(2) The recorded footage was passed through the Depth Anything [25] AI network, producing a sequence of grayscale depth maps. These maps represented the spatial relationships
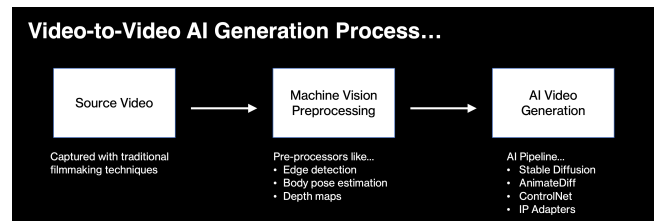

Figure 6: Diagram of video-to-video translation process using machine vision preprocessing. Note the video generation pipeline never "sees" the source video, only the intermediary pre-processed files.

Figure 7: Sample of macro-body part images of both artists used to power the IPAdapters, further blurring the boundaries between the entities.

of the wrestlers but struggled to delineate their entangled movements.

(3) These depth maps were fed into ControlNet, guiding the generative pipeline to produce video frames that maintained the wrestlers' ambiguous fluidity.

3.2 Content: IPAdapters

To create a coherent aesthetic, macro images of the artists' bodies were used as reference inputs for IPAdapters. These tools enable consistent style transfer and content preservation, guiding the generated video toward the intended "fleshy, non-representational" aesthetic. The macro imagery heightened the abstraction of the wrestlers' forms, contributing to the poetic ambiguity central to the work.

3.3 Clear Process for Blurred Boundaries

By integrating ControlNet's structural guidance with IPAdapters' stylistic conditioning, the artists facilitated a generative process that blurred bodily boundaries. This process was developed through a clear understanding of the technology's affordances and limitations. Rather than passively accepting AI's constraints, the artists shaped these limitations into the work's themes of intimacy and entanglement.

4 Discussion

4.1 Making Visible: Extending the Tradition of Experimental Filmmaking

Experimental filmmaking has long been a medium for exposing hidden dynamics, exploiting technological possibilities, and challenging dominant ways of seeing [22]. *Me vs. You* aligns itself with this tradition, drawing connections to two seminal practices: Eadweard Muybridge's motion studies and Kenneth Anger's queer cinema. Each of these works shares a commitment to making visible what might otherwise remain unseen, albeit through different methodologies and contexts.

4.1.1 Eadweard Muybridge—Visualizing Human Dynamics. One of the most direct precedents for *Me vs. You* is Eadweard Muybridge's early cinematic studies of male wrestlers, published in his seminal series Animal Locomotion [19]. Muybridge's experimental work captured sequential photographs of humans and animals in motion, isolating individual actions to analyze how bodies move through space and time. By fragmenting motion into discrete moments, Muybridge made visible the properties of physical movement that the human eye could not perceive unaided. Beyond its scientific

contributions, the work also hints at the interpretive nature of mechanical vision with its capacity to dissect, document, and classify. As Walter Benjamin [4] notes, technological systems like Muybridge's camera mediate and reconstruct human perception, an effect he described as the "optical unconscious."

Me vs. You relates to Muybridge's studies both aesthetically and conceptually. The AI-generated depth maps in *Me vs. You* share a strong visual resemblance to Muybridge's wrestler studies. Furthermore, both works make visible the underlying mechanics of their respective visual mediums, revealing how these technologies capture reality. However, where Muybridge's motion studies sought to standardize human movement, *Me vs. You* embraces the failure of AI systems to fully represent the nuance of human contact. The inability of the depth maps to resolve the wrestlers' entangled bodies into distinct, separable figures reflects a deeper tension between technological classification and the fluidity of human experience. This failure allows for a more ambiguous and poetic rendering of the physical encounter, transforming the limits of machine vision into an aesthetic strength. In this sense, *Me vs. You* critiques the legacy of mechanical representation embodied in Muybridge's work, while extending it into the realm of digital abstraction and generative filmmaking.
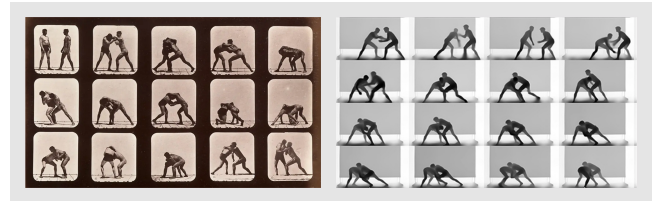

Figure 8: Left: Eadweard Muybridge's *Athletes. Wrestling* from 1881 [18] (Retrieved from the Library of Congress, https://www.loc.gov/item/2009630536/). Right: Sample of the *Me vs. You* depth maps from 2024. Early on in our experiments, we noticed a striking resemblance between the two.

4.1.2 Kenneth Anger—Visualizing Queer Identity. Both *Me vs. You* and Kenneth Angers's *Fireworks* [2] attempt to visualize queer experience through experimental film techniques. Anger's *Fireworks* is a landmark in queer experimental cinema, employing montage, symbolism, and a dreamlike aesthetic to explore the sadomasochistic interplay between aggression and sexuality. The film depicts a young man's violent yet erotically charged encounter with sailors, blending elements of fantasy and reality to articulate the complexities of queer desire. At the time, representations of homosexuality were largely censored in mainstream contexts [5]. Anger's use of experimental techniques allowed him to visualize ambivalent, contradictory aspects of gay identity that were otherwise repressed or invisible in mainstream media.

Similarly, *Me vs. You* uses generative AI to articulate the entangled dynamics of queer intimacy. The wrestling match, with its moments of physical struggle and closeness, becomes a site for exploring these contradictory dualities. Like Anger, we employ experimental techniques to render visible the complexities of desire

and intimacy that defy straightforward representation. The generative AI pipeline, with its inherent ambiguities, becomes a tool for articulating these dynamics in a new aesthetic language.

4.2 Subverting AI Limitations for Poetic Outcomes

Me vs. You embraces the limitations of machine vision systems as an artistic opportunity. Machine vision tasks such as classification, segmentation, and recognition are widely employed in generative AI processes [15, 21]. Kalluri et al. [14] describe how these computer vision tasks, such as depth maps, facial recognition, and pose detection, frequently end up in surveillance and military contexts. These tools demand precision and categorization, often reducing people into fragmented body parts and quantifiable data points. Kalluri et al. frame this phenomenon within the "Surveillance AI Pipeline," critiquing its reductive approach to human bodies, which prioritizes separation and classification to facilitate mass surveillance.

In *Me vs. You*, one such system is deliberately subverted to create poetic outcomes. We recognized that AI depth maps would struggle to separate the wrestlers' bodies when they became sufficiently intertwined. This "failure" produces the visual ambiguity at the heart of the project. Rather than serving as a tool for classification and control—as in surveillance contexts—the depth maps instead reflect the subjectivity and fluidity of human relationships. By exposing and recontextualizing this technological limitation, the project transforms the AI's failure into an artistic strength, illustrating a resonant view of intimacy that machine vision systems are inherently incapable of comprehending. This subversion critiques the hegemonic use of these technologies, repurposing them as tools for experimental and expressive ends.

5 Conclusion

From Muybridge's motion studies to the rise of digital video, image technologies have continuously shaped human perception, influencing both how we see and how we are seen. AI technologies extend this history in the era of deep learning, introducing new creative opportunities alongside significant critical challenges

Me vs. You demonstrates how experimental filmmaking is uniquely positioned to interrogate these challenges, making AI's mechanisms visible while subverting its limitations for expressive ends. Rather than treating AI as a tool for passive simulation, the work engages with the technology as a material—one with inherent affordances and constraints that shape its aesthetic and conceptual outcomes. This method draws on traditions of structuralist filmmaking [11] and active divergence [6], where exposing a medium's limitations becomes a generative act. *Me vs. You* adopts this technique to transform AI's failures of perception into a poetic exploration of queer intimacy, fluidity, and instability.

This approach exemplifies *Creativity for Change* in the context of AI and experimental media. By subverting the reductive logic of computer vision, the work challenges algorithmic frameworks of classification and control—opening up more sensual modes of representation and expansive ways of seeing.

References

[1] ACM Creativity and Cognition. 2024. Creativity for Change, Creativity & Cognition Conference 2025. https://cc.acm.org/2025/.
[2] Kenneth Anger. 1947. Fireworks.
[3] Francis Bacon. 1972. Triptych August 1972.
[4] Walter Benjamin. 1996. A Little History of Photography. In *Selected Writings: 1927-1934*, Howard Eiland and Gary Smith (Eds.). Harvard University Press.
[5] Benshoff and Griffin. 2005. *Queer Images: A History of Gay and Lesbian Film in America*. Rowman & Littlefield Publishers.
[6] Terence Broad, Sebastian Berns, Simon Colton, and Mick Grierson. 2021. Active Divergence with Generative Deep Learning – A Survey and Taxonomy. doi:10.48550/arXiv.2107.05599 arXiv:2107.05599 [cs]
[7] Judith Butler. 2011. *Gender Trouble: Feminism and the Subversion of Identity*. Routledge.
[8] Adam Cole, Gregor Petrikovič, and Mick Grierson. 2024. Me vs. You: On Queer Intimacy and the Interpretive Limits of AI Vision Systems. In *SIGGRAPH Asia 2024 Art Gallery (SA Art Gallery '24)*. Association for Computing Machinery, New York, NY, USA, 1. doi:10.1145/3680529.3688969
[9] Jonathan Crary. 1992. *Techniques of the Observer: On Vision and Modernity in the Nineteenth Century*. MIT Press.
[10] PLAN C Social Innovation Foundation. 2024. mAI future: The many faces of AI. https://maifuture.pl/.
[11] Malcolm Le Grice. 2001. *Experimental Cinema in the Digital Age*. Bloomsbury Publishing.
[12] Yuwei Guo, Ceyuan Yang, Anyi Rao, Zhengyang Liang, Yaohui Wang, Yu Qiao, Maneesh Agrawala, Dahua Lin, and Bo Dai. 2024. AnimateDiff: Animate Your Personalized Text-to-Image Diffusion Models without Specific Tuning. doi:10.48550/arXiv.2307.04725 arXiv:2307.04725 [cs]
[13] Oliver Hangl. 2024. K48: Model For Me. https://www.oliverhangl.com/projects/curating/k48/k48-2024/k48-189_HcMF.html.
[14] Pratyusha Ria Kalluri, William Agnew, Myra Cheng, Kentrell Owens, Luca Soldaini, and Abeba Birhane. 2023. The Surveillance AI Pipeline. doi:10.48550/arXiv.2309.15084 arXiv:2309.15084 [cs]
[15] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep Learning. *Nature* 521, 7553 (May 2015), 436–444. doi:10.1038/nature14539
[16] Lev Manovich. 2002. *The Language of New Media*. MIT Press.
[17] José Esteban Muñoz. 2009. *Cruising Utopia: The Then and There of Queer Futurity*. NYU Press.
[18] Eadweard Muybridge. 1881. Athletes. Wrestling.
[19] Eadweard Muybridge. 1887. *Animal Locomotion: An Electro-photographic Investigation of Consecutive Phases of Animal Movements*. University of Pennsylvania, Philadelphia.
[20] Eadweard Muybridge. 1901. Some Phases in a Wrestling Match. In *Human Figure in Motion*. London, 75.
[21] Alex Nichol, Prafulla Dhariwal, Aditya Ramesh, Pranav Shyam, Pamela Mishkin, Bob McGrew, Ilya Sutskever, and Mark Chen. 2021. GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models.
[22] A. L. Rees. 2011. *A History of Experimental Film and Video: From the Canonical Avant-garde to Contemporary British Practice*. Palgrave Macmillan.
[23] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis With Latent Diffusion Models (Stable Diffusion). In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10684–10695.
[24] Tate. 2025. Late at Tate Britain: 80s Valentines Special | Tate Britain. https://www.tate.org.uk/whats-on/tate-britain/the-80s-photographing-britain/late-at-tate-britain-80s-valentines-special.
[25] Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. 2024. Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10371–10381.
[26] Hu Ye, Jun Zhang, Sibo Liu, Xiao Han, and Wei Yang. 2023. IP-Adapter: Text Compatible Image Prompt Adapter for Text-to-Image Diffusion Models. doi:10.48550/arXiv.2308.06721 arXiv:2308.06721 [cs]
[27] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding Conditional Control to Text-to-Image Diffusion Models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 3836–3847.

A Installation Requirements

A.1 Physical Exhibition

A.1.1 Installation Requirements. The core installation requires two closely linked displays. The scale of the central installation is flexible and can be shrunk or expanded depending on the AV materials available and the scale of the space. Similarly, the orientation of the screens can be adjusted to one-on-top of the other instead of perpendicular.
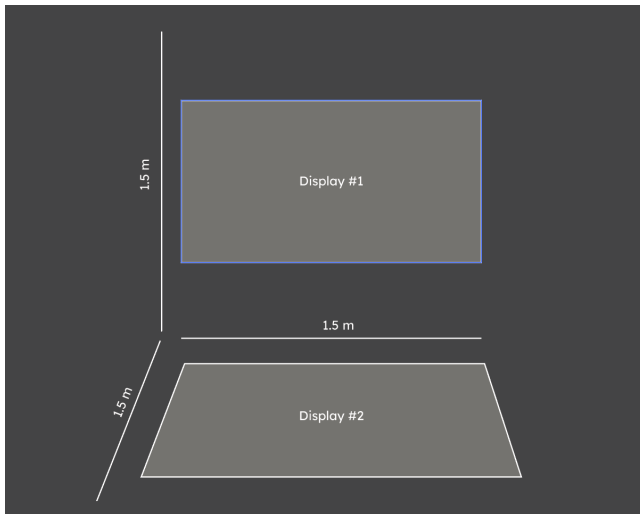

Figure 9: Ideal orientation of two-channel installation.

A.1.2 AV Requirements.
(1) Two high-quality televisions or projectors with HDMI input (55"+ preferred)
(2) Speakers or audience headphones
(3) Two synced media players or computer with dual 4K HDMI output

A.2 Online Exhibition

Two-channel video displayed in virtual space.

B Online Media Resources

(1) **Full dual-channel video**: https://youtu.be/YxY03DmtKwQ
(2) **Tate Britain installation**: https://youtu.be/_UWX_YqLC84
(3) **Project website**: https://www.adamcole.studio/work/me-vs-you