

Kiss/Crash: Using Diffusion Models to Explore Real Desire in the Shadow of Artificial Representations

ADAM COLE, Creative Computing Institute, University of the Arts London, United Kingdom

MICK GRIERSON, Creative Computing Institute, University of the Arts London, United Kingdom



Fig. 1. Installation of *Kiss/Crash*: a collection of three pieces that use AI image-translation to meditate on the nature of desire in the age of AI. On the central screen is the audio-visual piece *Kiss/Crash*, in which a car crash is turned into a kiss, projected on either side of the central screen is *Me Kissing Me*, a video in which the artist in a slow kiss with himself is transformed via AI, and central within the space is *Crash Me, Gently*, an interactive installation consisting of a television and foot pedal that allows viewers to accelerate the intensity of the kiss on the screen. (© 2022 Adam Cole and Hannah Burton)

This paper describes the art installation of *Kiss/Crash*: a multi-screen work exploring the subject of AI-imagery and representation as well as the autobiographical themes of loneliness, desire, and intimacy in the digital age. The installation consists of three individual works in a shared space, *Kiss/Crash*, *Me Kissing Me*, and *Crash Me, Gently*, which all play with augmenting, inverting, and negating the iconic image of the kiss using diffusion-based image translation. In the production of this work, several potentially novel video translation techniques were developed and refined to create high-quality results central to these pieces. Throughout our

Authors' addresses: Adam Cole, Creative Computing Institute, University of the Arts London, 45-65 Peckham Rd, London, United Kingdom, SE5 8UF, a.cole@arts.ac.uk; Mick Grierson, Creative Computing Institute, University of the Arts London, 45-65 Peckham Rd, London, United Kingdom, SE5 8UF, m.grierson@arts.ac.uk.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM.

2577-6193/2023/8-ART \$15.00

<https://doi.org/10.1145/3597625>

research, we reflect on the nature of images and place diffusion models within a history of image-production technologies that artists have been contending with and responding to for the past one hundred years. This paper aims to extend that artistic tradition with a provocative, original aesthetic and technique that reveals the logic of AI imagery and hints at how our relationship to reality will continue to be stretched and shaped by artificial representations at an accelerating pace.

CCS Concepts: • **Computing methodologies** → **Neural networks; Animation**; • **Applied computing** → **Media arts**.

Additional Key Words and Phrases: diffusion models, generative AI, image-to-image translation, video translation, fine art

ACM Reference Format:

Adam Cole and Mick Grierson. 2023. Kiss/Crash: Using Diffusion Models to Explore Real Desire in the Shadow of Artificial Representations. *Proc. ACM Comput. Graph. Interact. Tech.* 6, 3 (August 2023), 11 pages. <https://doi.org/10.1145/3597625>

1 INTRODUCTION

1.1 Motivation

The creation of *Kiss/Crash*, an audio-visual installation presented in Figure 1, was motivated by the recent proliferation of AI-generated content. In a world already drowning in images, the oncoming tidal wave of AI-generated “art” is reason for pause, if not outright concern. This is not just because of the ethical questions underlying the technology [Lee 2022], but because images are incredibly powerful: their ability to entertain, delight, and inspire is matched by their potential to misinform, control, and incite [Manghani et al. 2006].

The debate on the benefits and dangers of images has existed for hundreds of years but became especially vital in the post-industrial 20th century, when the production, distribution, and consumption of images became deeply tied to technological advancements like photography, mass media, and instant visual communications [Manghani et al. 2006].

This discourse can easily be extended into the 21st century, especially in relation to AI-generated images. For example, it is not a difficult mental leap to imagine how an image-generating machine fits neatly within the logic of late capitalism [Jameson 1992], exaggerates the experience of hyperreality [Baudrillard 1994], or amounts to another evolution of spectacle [Debord 2005]. However, one view of specific critical importance here is Susan Sontag’s *On Photography* where she describes our modern relationship to images as follows: “A society becomes modern when one of its chief activities is producing and consuming images, when images that have extraordinary powers to determine our demands upon reality, and are themselves coveted substitutes for firsthand experience, become indispensable to the health of the economy, the stability of the polity, and the pursuit of private happiness” [Sontag 2019].

If we are to accept that images have these extraordinary powers, how will a technology which can generate endless images in an instant impact our economy, polity, and pursuit of happiness? And, how then, are artists expected to use this technology responsibly?

While there is no right answer, we’d benefit greatly by recognizing that this technology is revolutionary in its capabilities but familiar in the questions it raises about the nature of images, representation, and authenticity. As such, we can look to past art movements which responded to these same questions for strategies in confronting the core issue. Relevant examples include Dada collage, Pop Art appropriation, and Postmodern pastiche. In all these cases, artists used the prevalent image-making technology of their time and *translated* those images in some way that both highlighted societal issues and revealed something unspoken about the way that technology worked.

1.2 Research Goals

Using these artists as inspiration, our research question is how the diffusion-based image translation process can be applied in a self-reflective way that reveals something true about AI-generated imagery and our digital culture in general. Additionally, how can this technique be used to address the artistic themes of particular interest to us, specifically interrogating the way intimacy is mediated through technology and the gap between real experience and its artificial representations?



Fig. 2. From left to right, an example of Dada [Hausmann 1919], Pop [Warhol 1962], and Postmodern [Kruger 1987] art which *translates* existing imagery into self-reflective, critical art. (© 2020 ADAGP, Paris and DACS, London, © 2018 The Andy Warhol Foundation for the Visual Arts, Inc. / Artists Right Society, New York and DACS, London, © 1987 Barbara Kruger)

2 RELATED WORKS

Diffusion-based models are a class of generative AI models that exploded in popularity after the release of text-to-image software such as DALL-E2, MidJourney, and Stable Diffusion [Marche 2022]. The generative diffusion process, as defined in [Ho et al. 2020] works by incrementally destroying a data sample through small additions of Gaussian noise and then training a model to learn how to reverse that noising process. The result is a model able to progressively move from complete noise toward a clean sample representative of the dataset.

Image translation introduced by [Meng et al. 2021] and further developed by [Rombach et al. 2021] allows us to modify an existing image by applying some degree of noise to the original and then denoising that image while guiding the result toward a text prompt as seen in Figure 3. For video translation, we break the input video into individual frames, run image translation on each input frame, and regroup the generated images into a new video.

Engineers and artists have experimented with this technique using open-source tools like DiscoDiffusion [Crowson 2021] to develop experimental video art [Rainisto 2022], cinematic special effects [@CorridorDigital 2022], and 3D environments [@ScottieFoxTTV 2022]. However, image translation has been a rich topic in generative-AI long before the recent focus on diffusion models with techniques such as style transfer [Gatys et al. 2015] and pix2pix [Isola et al. 2016]. Two relevant projects which use image translation to critically probe AI networks are *Auto-Encoding Bladerunner* [Broad and Grierson 2017], which exposes the imperfect way generative systems encode and represent “reality”, and *Learning to See* [Akten et al. 2019], which reveals the learned bias of AI networks and how that reflects our own self-affirming cognitive biases.

These two works are by no means an exhaustive review of critical art using AI image translation. Rather, these projects provide a good blueprint for merging technical research with the creation of a visual art that reflects on the very nature of the images generated. We will use a similar method



Fig. 3. Example of image-to-image translation with the prompt, “A fantasy landscape, trending on artstation”. Left is the original, and right is the result. [Rombach et al. 2021] (© 2022 Computer Vision LMU Munich)

in our research into diffusion models and their artistic potential, eventually arriving at the final production of *Kiss/Crash*.

3 PROCESS

To achieve the results in *Kiss/Crash*, we developed and refined several potentially novel video translation techniques over the course of several months through small targeted animation studies outlined in [Cole and Grierson 2022]. Some of these techniques concurrently emerged through various open-source tools including [Crowson 2021] and [Deforum 2022], which we adapted and combined to achieve the original, high-quality results seen in these works. While a detailed survey of the issues and solutions used in this process can be reviewed in the full technical report [Cole and Grierson 2022], the high-level challenges and discoveries are outlined here.

The difficulty with diffusion-based video translation is that while each generated frame needs to be a successful translation of the input frame, the video as a whole needs to maintain temporal and aesthetic coherence. This means each translated frame needs to relate to the original input as well as the other frames already generated.

One such technique to handle this problem is *frame-blending*, where the input frame to be translated is blended with the previously generated output. We then pass the blended input through the diffusion process, a trick that allows for more temporal coherence by relating the new image not just to the input frame but to the frame that will play before it. We also made use of *WarpFusion*, a technology developed by Alex Spirin [Spirin 2022], which extends *frame-blending* by also using optical flow maps to enhance temporal coherence across the generated video. Finally, we used *DreamBooth* [Ruiz et al. 2022], a tool that allowed us to fine-tune a Stable Diffusion model on a small set of iconic Hollywood kisses. This focused the stylistic and compositional results of the diffusion process leading to more temporal and aesthetic coherence. The combined results of these techniques can be seen in Figure 4.

In addition to making the video results smooth and coherent, we also developed some techniques to make the process of AI image translation legible in the results. One such technique was *strength-scheduling*: as each clip progresses, more diffusion steps are applied to each input frame, resulting in a video that incrementally transforms from the input video to the prompt over time, as seen in Figure 5. We also used *prompt-interpolation*, which gradually morphs the prompt used to guide the image translation, allowing the video to smoothly transition through various scenarios, as seen in Figure 9. Lastly, we post-processed the raw generated clips using *frame-interpolation* software [Reda et al. 2022], which helps smooth out the video and enables the works’ cinematic, slow-motion aesthetic.



Fig. 4. Three clips of a car crash being smoothly morphed into a kiss using diffusion image-to-image translation. *Frame-blending* and *DreamBooth* help keep the composition of the video consistent over time. *Strength-scheduling* makes the process of translation visible. (© 2022 Adam Cole)



Fig. 5. Example of *strength-scheduling*. Results looks more like the input at the start and increasingly match the prompt by the end. (© 2022 Adam Cole)

4 OUTPUTS

4.1 Installation: An Image World

The three works introduced in this paper, *Kiss/Crash*, *Me Kissing Me*, and *Crash Me, Gently*, were exhibited alongside each other in one immersive installation, as seen in Figure 1. The cacophony of screens, sounds, and repetitive kiss imagery mimics a world already saturated in images and increased both the intensity and provocation of the work.

4.2 Kiss/Crash

In this piece, playing on the central screen of Figure 1, a car crash is turned into a kiss using AI image translation. The initial video used is a test crash of a classic 1950's car [for Highway Safety 2014]. Over a minute in length, the cars repeatedly crash into each other. Each time the colliding metal transforms into faces pressed deeply together, and the sound of the crash smoothly transitions into a romantic song rich in Hollywood nostalgia. As the video progresses, the impacts crescendo in speed and intensity, and the image translations become more erotic, violent, and unhinged until finally, the tension is released, and the original slow-motion video of the crash plays with no translation applied.

4.3 Me Kissing Me

Me Kissing Me, playing on either side of *Kiss/Crash* in Figure 1, is a video of the artist engaged in a slow romantic kiss with a mirror image of himself on a desolate black background. The slow-motion



Fig. 6. Photographs from the exhibition of *Kiss/Crash*. The over-saturation of artificial images reinforces the themes of the piece. (© 2022 Adam Cole)

kiss is translated via AI-generation into augmented and perverted versions of the original scene, seamlessly transforming the artist and his double into various kissing couples, such as cinematic lovers, religious idols, political celebrities, and uncanny creatures. Beneath the AI-generated scenes, the underlying video of the artist kissing himself gently fades in and out, reminding the viewer of the artificial nature of the generated romance.

4.4 Crash Me, Gently

Crash Me, Gently, embedded within the space in Figure 1, is an interactive installation, this time transforming a kiss into a crash. Here a vintage television loops a video of the artist kissing himself, which transforms into a smooth Hollywood kiss whenever the faces meet. Viewers are invited to accelerate a foot pedal placed in front of the TV, which increases the intensity of the images, causing them to become more cinematic but also increasingly disjointed, pornographic, and violent. Contexts blend together as celebrities, creatures, and images of destruction take the place of the central lovers. The audio also accelerates, with sounds of metal crashing whenever the couple's lips meet. The image becomes increasingly destabilized until the pedal is released, after which the TV returns to its calm loop.

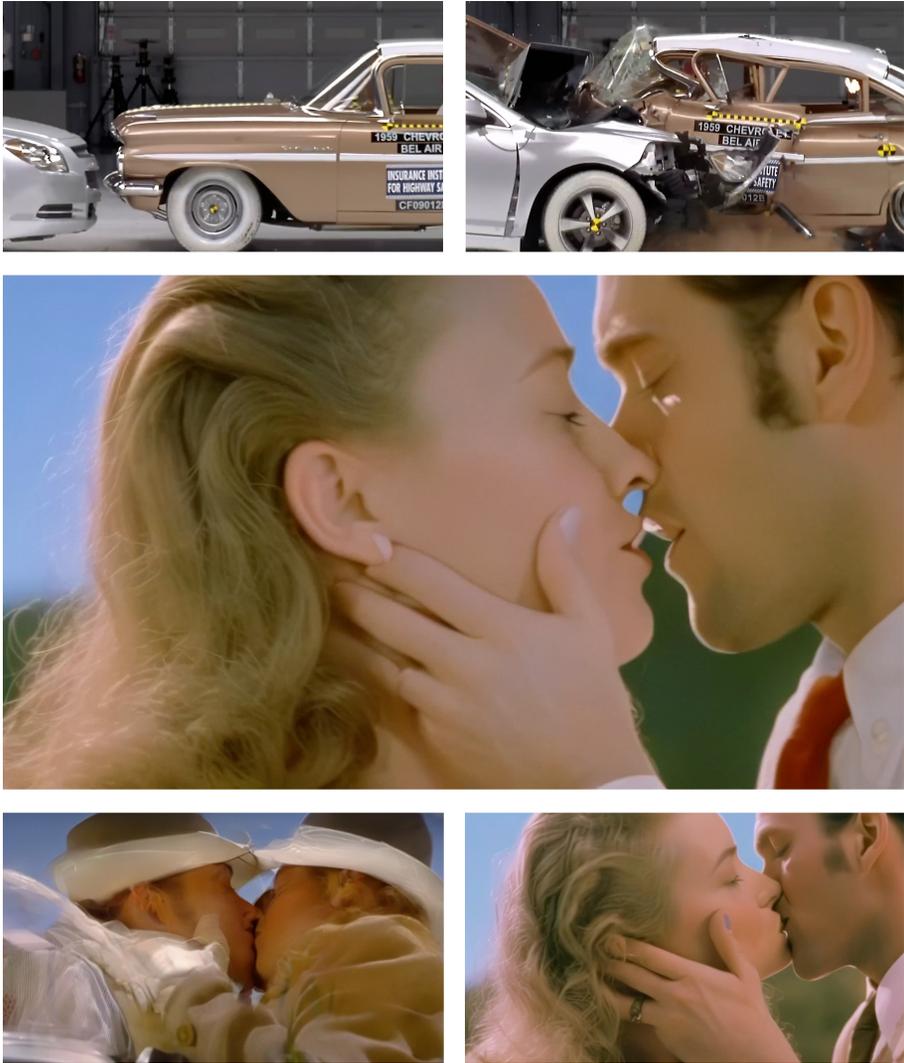


Fig. 7. Frames from the central *Kiss/Crash* video presented in Figure 1. First row is the input video untranslated, below are examples of the crash translated into a kiss at various stages. (© 2022 Adam Cole)



Fig. 8. Selected sequence from *Me Kissing Me* presented in Figure 1. Here, the artist in a kiss with himself fades in an out of the artificial scene. (© 2022 Adam Cole)



Fig. 9. Highlights from *Me Kissing Me*, as seen in Figure 1, where both the original input video and artificial layer on top of it are visible, exposing the process of AI image translation. *Prompt-interpolation* was used to smoothly transition through these scenes while maintaining a consistent, overall composition. (© 2022 Adam Cole)

5 DISCUSSION

In the following section, we will reflect on how these pieces fulfill this paper's stated goal of exploring the artistic potential of diffusion-based image translation. Specifically, we will reflect on how these pieces point to the gap between real experience and artificial representations within the context of AI and the autobiographical themes of desire and loneliness in the digital age.

5.1 Simulated Arousal

The central video of *Kiss/Crash* points to the power of images by creating a provocative metaphor between two seemingly unrelated but ubiquitous images from Western media: the car crash and the kiss. By appropriating its aesthetic from classic Hollywood, an industry that brands itself as *the dream factory*, this piece aims to relate AI-imagery to a history of image production technologies meant to incite our desires through artificial representations.

The anticipated experience of the viewer is one of arousal as they watch the crash repeatedly turned into an intimate scene, an experience reinforced by the acceleration of the video's rhythm and sexualized, violent content. But that desire, both in this work and in our culture in general, is fueled, homogenized, and repurposed by artificial representations. The generated desire is real.



Fig. 10. *Crash Me, Gently* interactive installation. At rest, the kissing loop on the screen is calm and gentle. When activated, the looping kiss is aggressive and extreme. (© 2022 Adam Cole)

The images are not. The audience is left to consider the way fictional representations in media, even ones that don't reflect the audience, can still arouse intense desires that can never be satisfied, and the way AI-generated content may continue that tradition.

5.2 Desire and Loneliness in Artificial Space

Both *Me Kissing Me* and *Crash Me, Gently*, highlight the contrast between the individual experience of loneliness and the desire for authentic, sensuous experiences with others. In both these pieces, the isolated central figure is in a cold mirrored kiss with himself, which is repeatedly transformed into a warm sensual scene. In these pieces, the self is quite literally projected into romanticized

representations, visually demonstrating the ways we subconsciously see ourselves in fictional images.

In *Me Kissing Me*, the process of image translation becomes visible as the original figures fade in and out of various generated poses, revealing the way AI tools can easily manipulate digital surfaces. These augmentations have the potential to be incredibly beautiful but can also feel quite shallow, like a thin, seductive veneer that obscures the identity of the original. In that sense, they become illustrative of our increasingly mediated relationship to reality in the digital era. One gets a sense of the way AI translation will stand between the original and our perception of it, especially with the specter of the metaverse on the horizon.

The central aspect of *Crash Me, Gently*, is its interactivity which invites the viewer to use the foot pedal to accelerate the intensity of the experience. The viewer becomes complicit in the hunt for more exaggerated representations of “love”. The frenetic changes of context as the pedal is pressed, including the fast-paced mixture of violence, sex, celebrity, and disaster, mimics the increasing schizophrenia of the digital world. The immediate result of pressing the pedal and being bombarded with these sensual scenes is perversely satisfying but also questions the way desire can be aroused by representations while never truly being satisfied.

All these pieces point to the underlying logic of AI-imagery by making the process of image translation visible and demonstrating the infinite layers of facades that may stand between us and the “authentic” ground truth. Together, they suggest a world where any spectacular desire can be explored, paving the way for a future where artificial representations stand in for real sensations at an accelerating pace.

6 CONCLUDING REMARKS

The objective of this paper was to repurpose a novel image-making technique with a critical eye in the same spirit of former avant-garde artists. The key contribution in this space was developing an art that makes the process of AI translation visible while hinting at its potential dangers. In the process of this development, we also explored several potentially novel techniques to create high-quality diffusion-based video translations that supported our aesthetic goal.

6.1 Looking Back to Look Forward

Susan Sontag, writing 50 years before the arrival of AI imagery, made the following call to action which seems just as relevant today, if not more so: “Images are more real than anyone could have supposed. And just because they are an unlimited resource, one that cannot be exhausted by consumerist waste, there is all the more reason to apply the conservationist remedy. If there can be a better way for the real world to include the one of images, it will require an ecology not only of real things but of images as well” [Sontag 2019].

We opened this essay with concern for the oncoming flood of artificial images, but our ultimate goal with this project was to consider the ways we can use artificial images in a self-reflective way that heightens our sense of the real. Our hope is that instead of creating more visual waste, we’ve added to the “worthwhile ecology of images”: an addition that doesn’t further drown us in the flood but instead helps us float to the surface where we can feel the sun on our skin. In doing so, we call on artists to reflect on how the AI tools we use in our practice today connect to the past, can be critical of the present, and might imagine a worthwhile future.

REFERENCES

Memo Akten, Rebecca Fiebrink, and Mick Grierson. 2019. Learning to See: You Are What You See. In *ACM SIGGRAPH 2019 Art Gallery* (Los Angeles, California) (SIGGRAPH '19). Association for Computing Machinery, New York, NY, USA, Article 13, 6 pages. <https://doi.org/10.1145/3306211.3320143>

- Jean Baudrillard. 1994. *Simulacra and simulation*. University of Michigan Press.
- Terence Broad and Mick Grierson. 2017. Autoencoding Blade Runner: Reconstructing Films with Artificial Neural Networks. In *ACM SIGGRAPH 2017 Art Gallery* (Los Angeles, California) (*SIGGRAPH '17*). Association for Computing Machinery, New York, NY, USA, 376–383. <https://doi.org/10.1145/3072940.3072964>
- Adam Cole and Mick Grierson. 2022. Old Sights, New Visions: Controlled Uses of Diffusion Based Image-to-Image Translation for Generative Video, Technical Report. (November 2022). <https://ualresearchonline.arts.ac.uk/id/eprint/20056>
- @CorridorDigital. 2022. The spider-verse joins the MCU. <https://twitter.com/CorridorDigital/status/1591958349612224514>
- Katherine Crowson. 2021. Disco diffusion. <http://discodiffusion.com/>
- Guy Debord. 2005. *The society of the spectacle*. Rebel Press.
- Deforum. 2022. Deforum Stable Diffusion. https://colab.research.google.com/github/deforum-art/deforum-stable-diffusion/blob/main/Deforum_Stable_Diffusion.ipynb
- Insurance Institute for Highway Safety. 2014. 1959 Chevrolet Bel Air vs. 2009 Chevrolet Malibu IIHS crash test. https://www.youtube.com/watch?v=C_r5UJrxck&t=1s
- Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. 2015. A Neural Algorithm of Artistic Style. arXiv:1508.06576
- Raoul Hausmann. 1919. The Art Critic. <https://www.tate.org.uk/art/artworks/hausmann-the-art-critic-t01918>
- Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising Diffusion Probabilistic Models. arXiv:2006.11239
- Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. 2016. Image-to-Image Translation with Conditional Adversarial Networks. arXiv:1611.07004
- Frederic Jameson. 1992. *Postmodernism or the cultural logic of late capitalism*. Duke Univ. Press.
- Barbara Kruger. 1987. Untitled (I shop therefore I am). <https://www.artsy.net/artwork/barbara-kruger-untitled-i-shop-therefore-i-am>
- Hye-Kyung Lee. 2022. Rethinking creativity: creative industries, AI and everyday creativity. *Media, Culture & Society* 44, 3 (2022), 601–612. <https://doi.org/10.1177/01634437221077009>
- Sunil Manghani, Arthur Piper, and Jon Simons. 2006. *Images: A Reader*. SAGE.
- Stephen Marche. 2022. We're witnessing the birth of a new artistic medium. <https://www.theatlantic.com/technology/archive/2022/09/ai-art-generators-future/671568/>
- Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. 2021. SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations. arXiv:2108.01073
- Roope Rainisto. 2022. Temporary. <https://www.youtube.com/watch?v=Ev2ddH5PJr0>
- Fitsum Reda, Janne Kontkanen, Eric Tabellion, Deqing Sun, Caroline Pantofaru, and Brian Curless. 2022. FILM: Frame Interpolation for Large Motion. arXiv:2202.04901
- Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2021. High-Resolution Image Synthesis with Latent Diffusion Models. arXiv:2112.10752
- Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. 2022. DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation. arXiv:2202.04901
- @ScottieFoxTTV. 2022. Stable diffusion in VR + touchdesigner = realtime immersive latent space. <https://twitter.com/ScottieFoxTTV/status/1578387866572525570>
- Susan Sontag. 2019. *On Photography*. Penguin books.
- Alex Spirin. 2022. Disco Diffusion v5.2 - Warp by Alex Spirin. https://colab.research.google.com/github/Sxela/DiscoDiffusion-Warp/blob/main/Disco_Diffusion_v5_2_Warp.ipynb
- Andy Warhol. 1962. Marilyn Diptych. <https://www.tate.org.uk/art/artworks/warhol-marilyn-diptych-t03093>

A LINKS TO VIDEO MATERIALS

Kiss/Crash Video of completed work <https://youtu.be/UXiSLu7mVHA>

Me Kissing Me Video of completed work <https://youtu.be/KCcJp7fV9ac>

Crash Me, Gently Documentation video of interactive installation <https://youtu.be/YE1gdVxbuws>