KI-Kunst als Skulptur

Fabian Offert

Abstract: *Dieser Aufsatz diskutiert die bildwissenschaftliche Einordnung der KI-Kunst. Er stellt die These auf, dass seit 2021 nicht mehr primär die Frage nach den bildnerischen Fähigkeiten der Maschine im Zentrum der KI-Kunst steht, sondern die nach den ästhetischen Grenzen der Kunst. Der Übergang von ausschließlich bildbasierten zu multimodalen Verfahren ist ein technisch induzierter Laokoon-Moment, in dem nicht mehr – mit Walter Benjamin gesprochen – das mimetische Vermögen eines Mediums von Interesse ist, sondern die Grenzen der Abbildbarkeit überhaupt verhandelt werden. Diese Abhängigkeit der ästhetischen von der technischen Entwicklung lässt sich wiederum nur vollständig rekonstruieren, wenn man KI-Kunst als Skulptur versteht, also als subtraktives, nicht additives, plastisches Verfahren.*

Sechs mit dem Stable-Diffusion-Modell generierte Bilder, der Prompt lautete wie folgt: »Still life with four bunches of grapes, an attempt at creating life-like grapes like those of the ancient painter Zeuxis, Juan El Labrador Fernandez, 1636, Prado, Madrid.«

© Fabian Offert

Einleitung

> »This release is the culmination of many hours of collective effort to create a single file that compresses the visual information of humanity into a few gigabytes.« Stability.ai Stable Diffusion release announcement[1]

Die Geschwindigkeit der technischen Entwicklung im Bereich der KI-gestützten Bildsynthese sei anhand der Genese dieses Textes illustriert. Der schmale Ausschnitt der technischen Welt, auf den sich der Text bezieht, wurde im Verlauf weniger Monate vom Kopf auf die Füße gestellt. Jene Verfahren, die Anfang des Jahres 2022 noch als künstlerische Experimente in Erscheinung traten, sind im Herbst 2022 (zum Zeitpunkt der letzten Revision dieses Textes) nicht nur von den großen KI-Unternehmen übernommen, adaptiert und kommerzialisiert, sondern bereits umgehend wieder *reverse engineered* und der Öffentlichkeit zur Verfügung gestellt worden. Die KI-gestützte Bildsynthese – und damit die KI-Kunst –, so lässt sich vermuten, steht eigentlich noch ganz am Anfang.

Der vorliegende Text ist der Versuch einer systematischen Aufarbeitung einer sich gerade erst ausdifferenzierenden ästhetischen Praxis. Dabei geht es im Besonderen um eine genauere Beschreibung der Bildsynthese in der KI-Kunst, als sie die Kunstgeschichte bisher zu leisten vermochte. Am Anfang dieses Textes steht demnach lediglich die pragmatische Feststellung, dass die KI-Kunst im Sinne Frieder Nakes (1971) eine hinreichende Anzahl neuartiger Methoden der Bildproduktion und -manipulation etabliert hat, die es rechtfertigen, sie als eine eigenständige ästhetische Domäne zu beschreiben. Es soll weniger die Frage diskutiert werden, wie genau die KI-Kunst insgesamt ästhetisch zu beurteilen sei – zum Beispiel, ob wir tatsächlich einer künstlerischen Revolution beiwohnen, wie es immer wieder von KI-Künstler*innen selbst und Akteuren des Kunstmarktes behauptet wird (vgl. Bogost 2019; Offert 2019). Vielmehr soll es darum gehen, wie die konkreten, in der KI-Kunst künstlerisch rekontextualisierten Techniken der Künstlichen Intelligenz ästhetisch einzuordnen sind.[2]

---

1 https://stability.ai/blog/stable-diffusion-public-release. Zugegriffen am 15. November 2022.

2 Dies schließt ein, dass auch die politisch-sozialen Aspekte der KI-Kunst, zum Beispiel ihre vielschichtige Beziehung zum Überwachungskapitalismus, hier zunächst keine Rolle spielen werden. Es sei in diesem Zusammenhang auf die Arbeiten und Analysen

Im Zentrum des Textes steht eine im weitesten Sinne historische These: Eine spezifische technische Entwicklung – der Wechsel von GAN-basierten hin zu auf Transformer-Verfahren beruhenden (multimodalen) Bildsyntheseverfahren – hat die ästhetische Entwicklung der KI-Kunst signifikant beeinflusst. Genereller: Im Zentrum der KI-Kunst steht seit 2021 nicht mehr primär die Frage nach den bildnerischen Fähigkeiten der Maschine, sondern die nach den ästhetischen Grenzen der Kunst. Der Übergang von ausschließlich bildbasierten zu multimodalen Verfahren ist ein technisch induzierter Laokoon-Moment, in dem nicht mehr – mit Walter Benjamin (1933) gesprochen – das mimetische Vermögen eines Mediums von Interesse ist, sondern die Grenzen der Abbildbarkeit überhaupt verhandelt werden. Diese Abhängigkeit der ästhetischen von der technischen Entwicklung – so die These des Beitrags – lässt sich wiederum nur vollständig rekonstruieren, wenn man KI-Kunst als Skulptur versteht, also als subtraktives, nicht additives, plastisches Verfahren.

Plastik und Skulptur in der frühen Computerkunst

Um die Sinnhaftigkeit dieser Differenzierung deutlich zu machen, muss zunächst herausgestellt werden, wie prävalent ›plastisches‹ und wie selten ›skulpturales‹ Arbeiten im Digitalen ist. Plastik und Skulptur sollen hier selbstredend im übertragenen Sinne verstanden werden. Plastik ist also nicht im Sinne Kittlers (1993) als Schichtung von Elektronen zu denken, sondern als Inbegriff einer additiven Herangehensweise, die einen ›leeren‹ digitalen Raum nach und nach mit digitalen Grundbestandteilen (*pixel*, *vertex*, *voxel* etc.) auffüllt. Komplexität wird schichtweise aufgebaut, entweder manuell oder algorithmisch.

Schon die frühe Computerkunst der 1950er und 1960er Jahre beginnt ausnahmslos *from scratch*. Die Gründe hierfür sind sowohl ästhetische als auch pragmatische. Zum einen ist im Kontext der Informationsästhetik mit Frieder Nake (1974) die generative Ästhetik von der analytischen zu unterscheiden: Während existierende Kunst den Entwurf eines ästhetisch-algorithmischen Systems inspirieren und beeinflussen kann, erfordert die Ausführung dieses Entwurfs einen schwarzen Bildschirm und/oder ein weißes Blatt Papier. Zum

Adam Harveys (etwa 2021) verwiesen, der sich in besonderem Maße künstlerisch mit den politisch-sozialen Aspekten des maschinellen Lernens auseinandergesetzt hat.

anderen entsteht die frühe Computerkunst maßgeblich als ästhetisches Verfahren für eine bestimmte Art von essenziell additiver Hardware, nämlich den sogenannten Stiftplotter. Frieder Nakes erste algorithmische Arbeiten sind bloße Testmuster, um die Funktionen eines selbstentwickelten Treibers für den ZUSE Graphomat Z 64 auf Herz und Nieren zu prüfen (vgl. Nake 2021).

Gleichzeitig aber etabliert bereits die frühe Computerkunst die besondere Rolle der Selektion. Sollen im Sinne Nakes singuläre ›Meisterwerke‹ durch künstlerisch-algorithmische *Systeme* abgelöst werden, entsteht die Notwendigkeit, aus einer Vielzahl potenzieller *Objekte* auszuwählen. Je zentraler das für die frühe Computerkunst essenzielle Zufallselement ist, desto größer ist der Möglichkeitsraum, der geschaffen wird. Wenn also dabei das Objekt hinter das System zurücktritt, das Besondere im Allgemeinen aufgeht, so ist es dennoch präsent als Ergebnis eines Selektionsprozesses, als Repräsentant eines Systems, das als solches nicht ausstellbar ist. Viel mehr als der zum Statthalter der künstlerischen Intuition erklärte Zufall ist deshalb die Selektion das Refugium intuitiver Entscheidungen in der frühen Computerkunst. Die Spannungen zwischen algorithmischem System und künstlerischer Autonomie, zwischen der Ausarbeitung aller Konsequenzen eines Prinzips (vgl. Turing 1950) und dem ästhetisch-kuratorischen Eingriff in diese Konsequenzen machen in vielen Werken den ästhetischen Mehrwert gerade aus und setzen sie von bloßen *tech demos* ab. Erste Ansätze eines grundsätzlich subtraktiven Verfahrens in Form einer diskreten Selektion aus möglichen Repräsentanten eines ästhetisch-algorithmischen Systems lassen sich also bereits in der frühen Computerkunst ausmachen.

Wo finden sich diese Ansätze nun in der zeitgenössischen KI-Kunst wieder? Zunächst: Wenn in diesem Text von KI-Kunst die Rede ist,[3] dann sind damit künstlerische Werke gemeint, in denen ausgewählte *Architekturen* neuronaler Netze und die aus ihnen hervorgehenden *Modelle* zum Mittel der Bildproduktion werden. Die Unterscheidung von Architektur und Modell, die für die frühe Computerkunst tautologisch wäre, ist für die zeitgenössische KI-Kunst von nicht zu unterschätzender Wichtigkeit. Während Architekturen als Forschungsergebnisse oft ohne Restriktionen der Fachöffentlichkeit zur Verfügung gestellt werden – sowohl theoretisch in akademischen Aufsätzen als auch praktisch als vorgefertigter, frei zugänglicher Open-Source-Computercode –, sind Modelle, also ›austrainierte‹ neuronale Netzwerke, unabhängig von ihrer

3 Für eine Systematisierung verschiedener Verständnisse von ›KI-Kunst‹ siehe Michael Klipphahn-Karges Beitrag in diesem Band.

Architektur oft proprietär. So sehr die offizielle Geschichte der Künstlichen Intelligenz die Bedeutung architektonischer Innovationen betont, so sehr muss die Geschichte der KI-Kunst eigentlich als Geschichte von Modellen gelesen werden. Diese verläuft jedoch weit weniger linear und ist oft durch sprunghafte Entwicklungen gekennzeichnet. Wenn also im Folgenden von bestimmten Architekturen die Rede ist – konkret von *generative adversarial networks* (Goodfellow et al. 2014; Karras et al. 2018, 2019), *vision transformers* (Esser et al. 2021) und *diffusion models* (Dhariwal et al. 2021) –, muss immer mitgedacht werden, dass konkrete Werke der KI-Kunst ausschließlich von konkreten Modellen abhängen.

Fünf Jahre GANs: 2015–2020

Generative adversarial networks (GANs) wurden um das Jahr 2014 herum zuerst von Ian Goodfellow als neuartige Architektur für neuronale Netzwerke vorgeschlagen (Goodfellow et al. 2014). Sie ergänzen Funktionsweisen des CNN, des *(deep) convolutional neural networks*, um einen Ansatz aus der Spieltheorie: Zwei voneinander unabhängige Systeme, ein ›klassisches‹, klassifizierendes CNN und ein ›invertiertes‹, also bildproduzierendes CNN, werden in einen Wettbewerb zueinander gestellt. Das bildproduzierende Netzwerk, der *generator*, hat Zugriff auf einen kontinuierlichen, hochdimensionalen (z.B. 100-dimensionalen) Vektorraum, den sogenannten *latent space*. Seine Grundfunktion ist es, aus jedem Punkt in diesem Raum (also jedem 100-dimensionalen Vektor) ein Bild generieren zu können. Zu Beginn des Trainingsprozesses sind die so entstehenden Bilder reine Zufallsbilder, das heißt zufällige Anordnungen von Pixeln. Diesem Netzwerk gegenüber steht das zweite, klassifizierende Netzwerk, das *discriminator* genannt wird. Dieses Netzwerk hat Zugriff auf die vom *generator* produzierten Bilder wie auch auf einen Trainingskorpus von ›realen‹ Bildern. In jeder Iteration des Trainingsprozesses wird dem *discriminator* nun ein Bild gezeigt (tatsächlich funktioniert das Training in sogenannten *mini batches*, das heißt, mehrere Bilder werden zu einem einzigen verkettet, um die Verarbeitung effizienter zu machen). Dessen Aufgabe ist es, zu entscheiden, ob das Bild aus dem Trainingskorpus realer Bilder stammt oder vom *generator* produziert wurde.

Was zunächst wie eine einfach zu lösende Aufgabe klingt – schließlich zeigen die Bilder des Trainingskorpus reale Motive, wohingegen die vom *generator* produzierten Bilder reine Zufallsbilder sind –, ist für den *discriminator*

zu Beginn des Trainingsprozesses tatsächlich schwierig zu entscheiden: So, wie der *generator* nicht ›weiß‹, wie realistische Bilder zu erzeugen sind, ›weiß‹ der *discriminator* nicht, was reale Bilder von Zufallsbildern unterscheidet. Beide Netzwerke beginnen also ›bei null‹ und – dies ist die Innovation der GAN-Architektur – lernen ›gemeinsam‹, ihre jeweiligen Fähigkeiten zu verbessern. Der *generator* wird besser darin, realistische Bilder zu erzeugen, und der *discriminator* wird besser darin, reale Bilder von Produkten des *generators* zu unterscheiden. Wenn man dabei die Balance der Fähigkeiten über einen langen Zeitraum bzw. über viele Trainingszyklen hinweg aufrechterhält, steht am Ende ein *generator*, der Bilder erzeugen kann, die aussehen, als gehörten sie zum Korpus der realen Bilder. Diese Bilder sind jedoch gerade keine Kopien der realen Bilder im Trainingskorpus – der *generator* kann schließlich gar nicht wissen, wie diese genau aussehen, da er keinen Zugriff auf das Trainingskorpus hat und Informationen über die Qualität seiner Erzeugnisse lediglich vom *discriminator* erhält. Stattdessen produziert er Bilder, die ähnliche Eigenschaften wie jene im Trainingskorpus aufweisen: Sie ähneln in Inhalt und Form zwar den realen Bildern, die dem GAN zur Verfügung gestellt wurden, sind aber in jeder Hinsicht neue Bilder. Dies bedeutet, dass am Ende des Trainingsprozesses ein *generator* steht, der aus jedem beliebigen Punkt eines *latent space* ein realistisches Bild erzeugen kann. Im Umkehrschluss bedeutet dies, dass wir – vermittelt durch die Fähigkeiten des *generators* – einen hochdimensionalen Vektorraum erhalten, der Milliarden von möglichen Bildern enthält.

Konkret als ästhetische Werkzeuge wahrgenommen wurden GANs zunächst als wichtige Komponente in sogenannten *Style-transfer*-Algorithmen, das heißt in KI-Systemen, die bestimmte formale Eigenschaften (zum Beispiel, um ein beliebtes Beispiel aus der technischen Literatur zu zitieren, die charakteristischen Farben und den charakteristischen Pinselstrich van Goghs) aus einem Bild extrahieren und auf ein anderes übertragen können. Zu diesem Zeitpunkt waren GANs als eigenständige generative Systeme jedoch noch notorisch instabil. Zentral war das Problem des *mode collapse*: die Entstehung eines *generators*, der nur eine sehr kleine Anzahl möglicher Bilder erzeugen kann. *Mode collapse* entsteht genau dann, wenn der *generator* es zu schnell schafft, den *discriminator* von seinen Kreationen zu ›überzeugen‹. Einige frühe Beispiele, etwa die von Radford et al. (2015) erzeugten »imaginären Schlafzimmer«, zeigten jedoch bereits damals, wozu GANs potenziell in der Lage waren.

Die tatsächliche künstlerische Erforschung von GANs beginnt daher erst nach einer Reihe von technischen Verbesserungen (Goodfellow et al. 2016). Erste Experimente wie Mike Tykas *Portraits of Imaginary People* (2017) nutzten vor allem die DCGAN-Architektur (Radford et al. 2015). Um das Problem der niedrigen Auflösung, mit dem frühe GAN-Architekturen ebenfalls zu kämpfen hatten, zu umgehen, entwickelte Tyka einen Prozess, der *GAN-sampling* und *up-sampling*, also die künstliche Erhöhung der Bildauflösung, miteinander verband.

Inwiefern kann dieses Verfahren aber nun als subtraktives beschrieben werden? Schließlich beginnen beide GAN-Subsysteme, wie gesagt, ›bei null‹, also ohne analytische (*discriminator*) oder synthetische (*generator*) Fähigkeiten. Vergleichen wir den untrainierten *latent space* eines GANs mit einem weißen Blatt Papier oder einem schwarzen Bildschirm, so wird deutlich: Im Gegensatz zum ›leeren‹ Medium ist der *latent space* bereits mit Bildern gefüllt. Genauer: Jeder spezifische Punkt im *latent space* entspricht bereits einem spezifischen Bild, noch bevor der *discriminator* auch nur ein einziges ›reales‹ Bild ›gesehen‹ hat. Mit anderen Worten, das Medium selbst, die Architektur des neuronalen Netzes, birgt bereits das volle Potenzial eines scheinbar unendlichen Bildraumes. Allerdings haben die zu diesem Zeitpunkt im Bildraum verfügbaren Bilder keinerlei repräsentativen Charakter: Sie stellen nichts dar, verweisen weder auf ein allgemeines, universelles noch auf ein partikulares Objekt. Dies ändert indes nichts an der Tatsache, dass sie Bilder sind, die existieren und auf einfache Art und Weise zum Vorschein gebracht werden können.

Auch dieses ›Zum-Vorschein-Bringen‹ muss hier kurz technisch kontextualisiert werden. Der *generator* eines GANs (der oft schon allein für sich als GAN bezeichnet wird, obschon er, wie oben beschrieben, lediglich ein Subsystem desselben ist) erzeugt ein spezifisches Bild aus einem spezifischen Punkt im *latent space*. Im Falle eines 100-dimensionalen Vektorraumes kann dieser Punkt durch 100 einzelne Werte beschrieben werden. Wie in einem euklidischen, zweidimensionalen Koordinatensystem zwei Werte ausreichen, um einen exakten Punkt auf einer Fläche anzugeben, so definieren 100 Werte einen eindeutigen Punkt in einem 100-dimensionalen Raum. Gehen wir von einem normalisierten Raum aus, so liegen diese Werte zwischen –1 und 1. Dies bedeutet aber keineswegs, dass die Anzahl möglicher Bilder auf 3^{100} begrenzt ist. Denn der *latent space* eines GANs ist ein kontinuierlicher Raum, das heißt, Koordinaten werden als Gleitkommazahlen angegeben, in ihrer Präzision nur begrenzt durch den eingesetzten Variablentypus. Gehen wir von einem GAN mit einer Präzision von 32 Bit (*single-precision floating-point format*) und einem 100-

dimensionalen *latent space* aus, so können theoretisch $(3.4028235\times10^{38})^{100}$ Bilder erzeugt werden. Praktisch ist die Anzahl begrenzt durch die Art des Samplings, durch Besonderheiten diverser Architekturen und insbesondere durch die Effizienz des Trainings.

Was hier allerdings – völlig unabhängig von der genauen, technisch bedingten Anzahl möglicher Bilder – offensichtlich wird: Die Herausforderung bei der künstlerischen Arbeit mit GANs besteht nicht in der Bilderzeugung, sondern im Finden von Mitteln und Wegen, ›interessante‹ Punkte im *latent space* zu identifizieren. Aus einer Vielzahl von Möglichkeiten, aus einem Überfluss an Material werden konkrete Artefakte in einem solchen Maße herausgearbeitet, dass sie das Potenzial des Materials transzendieren. KI-Kunst, so könnte deshalb polemischer formuliert werden, ist Skulptur, nicht Plastik.

Der klassische Witz über den Bildhauer, der gefragt wird, woher er denn wisse, dass eine bestimmte Statue sich in einem bestimmten Marmorblock versteckt gehalten habe, zieht seine Pointe aus der Unsichtbarkeit der eigentlichen künstlerischen Arbeit. Freilich ist das Werk in gewisser Hinsicht im Material angelegt, aber herausgearbeitet werden kann es eben nur von wenigen künstlerisch und technisch ausgebildeten Menschen, die einen besonderen historischen, methodischen und ästhetischen Zugriff auf ihre Lebenswelt haben und über das Vermögen verfügen, diesen Zugriff in die Arbeit am Material zu übersetzen. In der Welt der GANs jedoch ›weiß‹ nur die Maschine selbst, was sich hinter einzelnen Punkten im *latent space* verbirgt. Im besten Falle existiert eine halbwegs kohärente räumliche Verteilung, die bestimmte Bildbestandteile bestimmten Punkten im *latent space* zuordnet (die Informatik spricht hier von *disentangled representations*). Selbst diese ist dem *latent space* jedoch in keinem Falle ›anzusehen‹, sondern nur experimentell ermittelbar.

Die *manuelle* Erkundung des *latent space* kann als Essenz der KI-Kunst zwischen circa 2015[4] und 2020 betrachtet werden. Werke aus dieser Zeit finden dabei unterschiedliche Ansätze, dessen unfassbare Ausmaße zu thematisieren. Neben dem einfachen kuratierten *sampling* setzte sich dabei der sogenannte *latent space walk* als Format durch. Dessen Reiz liegt in der Flüssigkeit der Bewegung: Bilder, die nahe beieinander liegenden Punkten im *latent space* zugeordnet sind, unterscheiden sich auch visuell nur minimal, sodass eine Inter-

4 An anderer Stelle (Offert 2022) habe ich die zentrale Rolle herausgearbeitet, die dem 2015 veröffentlichten DeepDream-Algorithmus beim Aufkommen der KI-Kunst zukam.

polation möglich wird, in der sich Einzelbilder organisch ineinander aufzulösen scheinen. *Latent space walks* wurden zentrale Werkzeuge im Repertoire von (mittlerweile etablierten) Künstlern wie Memo Akten und Mario Klingemann. Auch Anna Ridlers vielzitiertes Schlüsselwerk *Mosaic Virus* (2018) besteht aus einem solchen *latent space walk*, in dem semantisch eigentlich abgeschlossene Bildbestandteile verschwimmen, sich verschieben, aufteilen und verdoppeln.

Wie schon in der frühen Computerkunst geht es also die Ausstellung eines ästhetischen Systems als ästhetisches Objekt. Die ästhetische Spannung zwischen System und Objekt ist das erste wichtige Merkmal der KI-Kunst der GAN-Epoche. Das Format des *latent space walks* erfüllt jedoch darüber hinaus eine pragmatische Funktion, die auf das zweite wichtige Merkmal der KI-Kunst der GAN-Epoche verweist. *Latent space samples* sind in fast allen Fällen des künstlerischen Einsatzes von GANs *nicht* fotorealistische Bilder. Während populäre Architekturen wie StyleGAN 2 (Karras et al. 2019) für ihre nahezu fotorealistischen Ergebnisse angepriesen wurden, ist dieser Fotorealismus nicht haltbar, wenn kleinere, individuellere und weniger homogene Datensätze zum Einsatz kommen. Kann zum Beispiel ein von Nvidia im Rahmen der StyleGAN-2-Architektur veröffentlichtes Modell, das anhand eines besonders homogenen Datensatzes von Gesichtern[5] (Flickr-Faces FFHQ) trainiert wurde, auch nur diese, nämlich fotorealistische Gesichter erzeugen, bleiben künstlerische Anwendungen, so sie denn über die bloße Reproduktion und Kritik dieser und ähnlicher Modelle hinausgehen und eigene Datensätze verwenden, auf produktive Weise hinter diesem Fotorealismus zurück.

Philipp Schmitts und Steffen Weiss' Serie von GAN-generierten Stühlen (2018)[6] ist ein gutes Beispiel für KI-Kunst, die deren *glitch*-Aspekt bewusst betont. Erst signifikante Eingriffe in den generativen Prozess führen hier zu einem ästhetischen Artefakt, das als reales Objekt überhaupt bestehen kann. Sofia Crespo verbindet in ihrer Arbeit *Neural Zoo* (2018) handkuratierte Datensätze mit den traditionellen Techniken Collage, Pastiche und Cyanotypie. Durch die ›analoge Weiterverarbeitung‹ digitaler Bilder werden dabei nicht nur die klassischen ›Fehler‹ neuronaler Netzwerke wie eine zu große Anzahl hoher Bildfrequenzen (Geirhos et al. 2019) ausgeglichen, sondern auch beständige Artefakte generiert, die der Flüchtigkeit und Beliebigkeit der GAN-

5 Zur problematischen Verbindung von Gesichtserkennung und Künstlicher Intelligenz siehe Meyer (2021).

6 https://steffen-weiss.design/the-chair-project-generating-a-classic. Zugegriffen am 15. November 2022.

Bildermengen entgegenstehen. In den Arbeiten von Sarah Rosalena Brady schließlich, die GAN-generierte Bilder als maschinengewebte Tapisserien ausgeben lässt, löst sich der naturalistische Charakter synthetischer Bilder vollends in ein formales Prinzip auf. Bradys Werk *Untitled* (2020) zeigt dementsprechend GAN-halluzinierte Planeten. Aus der Übersetzung von Pixeln in Maschen entsteht so das Bild eines die *fabric of reality* webenden neuronalen Netzwerks.

Mit Aaron Hertzmann (2020) ist es die *visual indeterminacy* dieser Bilder, ihre bildinhaltliche Unschärfe, die einen ästhetischen Effekt hervorbringt. Mit Brecht (1957) könnte man gar von einem technischen Verfremdungseffekt sprechen, der in der KI-Kunst zum Einsatz kommt. Die ästhetische Rahmung unvermeidbarer technischer Artefakte schafft eine Distanz zwischen Rezipient und Werk, die dessen kritische Reflexion begünstigt. Dieser Verfremdungseffekt, der in letzter Instanz auf die kuratierte Unvollkommenheit der generierten Bilder zurückfällt, ist das zweite wichtige Merkmal der KI-Kunst dieser Epoche.

Transformer und die Automatisierung der künstlerischen Selektion

Die GAN-Epoche der KI-Kunst endet abrupt. Wie Helena Sarin, eine der prominentesten Vertreter*innen der GAN-basierten KI-Kunst, am 27. August 2022 auf Twitter schreibt:

> Come to think of it, this exhibit [die Online-Ausstellung *Reflections in the Water*[7]; Anm. F. O.] […] was a perfect closure for the GAN period of AI art; I mean there was some solid GAN art created after and maybe still is created but that was the end of the movement as we knew it, the thrill is gone.[8]

Die im September 2021 eröffnete Ausstellung, auf die der Tweet verweist, versammelt Künstler*innen wie Mario Klingemann, Anna Ridler, Helena Sarin, Jake Elwes und viele andere Vertreter*innen der GAN-Epoche. Dass sie in Sarins Tweet als Abschlusspunkt verstanden wird, verweist auf einen radikalen

7 https://feralfile.com/exhibitions/reflections-in-the-water-9ov. Zugegriffen am 15. November 2022.

8 https://twitter.com/NeuralBricolage/status/1563645089288753153. Zugegriffen am 15. November 2022.

Paradigmenwechsel in der KI-Forschung und – in der Folge – in der KI-Kunst, der bereits 2021 begann.

Im Januar 2021 veröffentlichte OpenAI, das wohl bekannteste unter den nichtakademischen KI-Forschungsinstituten, sowohl eine neuartige Architektur als auch ein fertiges Modell namens CLIP (Radford et al. 2021). CLIP (die Abkürzung steht für *contrastive language-image pre-training*) verfolgt einen multimodalen Ansatz, das heißt, CLIP-Modelle bringen Text und Bilder in Zusammenhang. Möglich wird dies durch die Analyse von Bildern im Kontext, wie sie das Internet in hoher Anzahl bereithält: Jedes Bild, das auf einer Website erscheint, ist gemeinhin von Text umgeben, der in direktem Zusammenhang mit dem Bild steht. CLIP produziert aus diesen Daten einen gemeinsamen *latent space* für Text und Bild sowie entsprechende *encoder*, die ungesehene Bilder und ungelesenen Text in Punkte in diesem *latent space* übersetzen können. Eine konkrete praktische Anwendung ist dementsprechend die Herstellung von Bildbeschreibungen: ›Zeigt‹ man CLIP ein Bild, so ist es in der Lage, die dem Bild am besten entsprechenden Textpunkte im *latent space* auszugeben. CLIP hat jedoch selbst keine Bildsynthesefähigkeiten, sodass der umgekehrte Weg zunächst verschlossen blieb.

Dennoch kann das CLIP-Modell zur Bildsynthese eingesetzt werden, wenn man es an ein existierendes generatives Modell ›ankoppelt‹. Genauer: wenn man, statt den *latent space* eines Modells auf zufälligem Wege zu durchlaufen, an jedem Punkt prüft, ob das vom generativen Modell produzierte und von CLIP klassifizierte Bild näher an die angepeilten Textpunkte im *latent space* von CLIP gerückt ist. In den verbleibenden Monaten des Jahres 2021 konnte sich auf der Basis dieser einfachen Idee eine höchst vitale ›Szene‹ etablieren, die zahlreiche Varianten und Verbesserungen der durch CLIP gesteuerten Bildsynthese auf den Weg brachte.[9] Programmierer:innen und Künstler*innen wie Katherine Crowson und Ryan Murdoch veröffentlichten ihre Programme im Google CoLab-Format, das heißt als im Browser ausführbare, cloud-basierte und interaktive Skripte. Betrachtet man die Ergebnisse dieser vielfältigen Ansätze, wird offensichtlich, dass der Einsatz von CLIP jenes grundsätzliche Problem der vorhergehenden Epoche der KI-Kunst löst, das für den künstlerischen Einsatz zentral ist: Wie können visuell interessante Punkte im *latent*

9 Siehe https://ml.berkeley.edu/blog/posts/clip-art/ (zugegriffen: 15. November 2022) für einen gewissenhaften Überblick und Underwood (2021) für eine weiter gefasste kulturelle Einordnung des Prinzips latent space.

space gezielt angesteuert werden? Mit CLIP funktioniert dies gewissermaßen ›auf Zuruf‹, durch die gezielte Manipulation des eingegebenen Textes.

Den technischen Durchbruch brachte dabei die Kombination von CLIP mit sogenannten *diffusion models*, einer Klasse generativer neuronaler Netzwerke, die das *generator-discriminator*-Prinzip der GAN-Architektur durch das Prinzip der gelernten Kompression ersetzt. Im Frühjahr 2022 veröffentlichten OpenAI und Google eine ganze Reihe von Modellen, die dieses bisher nur als künstlerisch-technisches Experiment vorliegende Prinzip aufnahmen und kommerzialisierten. Beginnend mit DALL-E 2, das durch eine umfassende PR-Kampagne und künstliche Verfügbarkeitsbeschränkungen[10] einer weiten Öffentlichkeit bekannt gemacht wurde, begann die KI-gestützte Bildsynthese, die lange gerade mit der *Unvollkommenheit* synthetisierter Bilder identifiziert wurde, als Möglichkeit der fotorealistischen Bilderzeugung wahrgenommen zu werden. Den vorläufigen Abschluss – zum Zeitpunkt der letzten Revision dieses Textes – bildet Stable Diffusion[11] (Rombach et al. 2022), ein von einem großen Kollektiv von Forschenden und Akteur*innen der KI-Industrie veröffentlichtes Open-Source-Modell, das es im Hinblick auf die Qualität der synthetisierten Bilder durchaus mit DALL-E 2 aufnehmen kann.

Der Fokus der KI-Kunst verschiebt sich im Kontext dieses Paradigmenwechsels hin zu multimodalen Modellen, radikal: *prompt engineering*, also die gezielte Komposition von Texteingaben zur Erzeugung ganz bestimmter Bildinhalte, ersetzt die formalen Experimente der GAN-Epoche. Dennoch sind es keinesfalls nur semantische Aspekte des Bildes, die sich über *prompts* gezielt beeinflussen lassen. Durch die Einstreuung von bestimmten Stichwörtern lassen sich ebenso bestimmte ›Stile‹ provozieren. Die Verwendung des Stichworts ›Unreal Engine‹ sorgt für einen hyperrealistischen Stil, der sich an bekannten 3-D-Programmen (wie eben auch Unreal, eine von Epic Games entwickelte Computerspiel-Engine) orientiert. Das Stichwort ›trending on ArtStation‹ sorgt für einen ›populären‹ Stil, wie er von vielen Hobbykünstler:innen auf der

10 Das Web-Interface für DALL-E 2 (das Modell selbst bleibt weiterhin proprietär) war lange Zeit nur auf Anfrage und für Forschende und Akteure der KI-Industrie zugänglich. Im August 2022 wurde es schließlich auch einer breiteren Öffentlichkeit zur Verfügung gestellt und gleichzeitig kommerzialisiert: Für jedes generierte Bild muss ein bestimmter Betrag entrichtet werden.

11 Der folgende Twitter-Thread bietet einen guten Überblick über die Funktionsweise von Stable Diffusion: https://twitter.com/ai__pub/status/1561362542487695360. Zugegriffen: 15. November 2022.

Website ArtStation gepflegt wird. Die Angabe von bestimmten analogen Filmtypen (›Provia, Velvia, Fujifilm Superia‹) produziert fotorealistische Bilder, deren Spektrum das Profil der genannten Medien erstaunlich genau abbildet.

Damit kehren wir zurück zur Ausgangsfrage dieses Aufsatzes: Wie sind die konkreten Techniken der KI-Kunst ästhetisch einzuordnen? Mit CLIP, so könnte man sagen, kommt das der KI-Kunst immer schon inhärente skulpturale Moment zu sich selbst. *Latent spaces* können nun gezielt befragt, statt bloß passiv durchschritten werden. Damit tritt aber auch die KI-Kunst in eine neue Phase ein und entfernt sich weiter von traditionellen algorithmisch-künstlerischen Methoden. Statt – wie die frühe Computerkunst – platonische Grundformen durchzuexerzieren, erzeugt die KI-Kunst post CLIP gezielt Neues aus *found footage:* visuelle *musique concrète.* Die Skulpturen, die sie schafft, sind Monumente für das Internet, die Mutter aller *datasets*.

Literatur

Bogost, Ian. 2019. The AI-Art Gold Rush Is Here. *The Atlantic*. https://www.theatlantic.com/technology/archive/2019/03/ai-created-art-invades-chelsea-gallery-scene/584134/. Zugegriffen am 1. Feburar 2023.

Brecht, Bertolt. 1957. Die Straßenszene. Grundmodell einer Szene des epischen Theaters. In *Schriften zum Theater*. Frankfurt a.M.: Suhrkamp.

Dhariwal, Prafulla und Alex Nichol. 2021. Diffusion Models Beat GANs on Image Synthesis. https://doi.org/10.48550/arXiv.2105.05233.

Esser, Patrick, Robin Rombach und Björn Ommer. 2021. Taming Transformers for High-Resolution Image Synthesis. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR46437.2021.01268.

Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville und Yoshua Bengio. 2014. Generative Adversarial Nets. *Advances in Neural Information Processing Systems*. https://dl.acm.org/doi/10.5555/2969033.2969125.

Harvey, Adam und Jules LaPlace. 2021. Researchers Gone Wild: Origins and Endpoints of Image Training Datasets Created »In the Wild«. In *Practicing Sovereignty: Digital Involvement in Times of Crisis,* Hg. von Bianca Herlo, Daniel Irrgang, Gesche Jost und Andreas Unteidig, 289–310. Bielefeld: transcript.

Karras, Tero, Samuli Laine und Timo Aila. 2018. A Style-Based Generator Architecture for Generative Adversarial Networks. https://doi.org/10.48550/arXiv.1812.04948.

Karras, Tero, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen und Timo Aila. 2019. Analyzing and Improving the Image Quality of StyleGAN. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/CVPR42600.2020.00813.

Meyer, Roland. 2021. *Gesichtserkennung*. Berlin: Wagenbach.

Nake, Frieder. 1971. There Should Be No Computer Art. *Bulletin of the Computer Arts Society*: 18–19. https://dam.org/museum/essays_ui/essays/there-should-be-no-computer-art/. Zugegriffen: 15. November 2022.

Nake, Frieder. 1974. *Ästhetik als Informationsverarbeitung. Grundlagen und Anwendung der Informatik im Bereich ästhetischer Produktion und Kritik*. Wien und New York: Springer.

Nake, Frieder. 2021. The Art of Being Precise. Interview mit Margit Rosen. https://www.youtube.com/watch?v=Z_pOiHX6HYE. Zugegriffen: 15. November 2022.

Offert, Fabian. 2019. The Past, Present, and Future of AI Art. *The Gradient*. https://thegradient.pub/the-past-present-and-future-of-ai-art/. Zugegriffen: 15. November 2022.

Offert, Fabian. 2023. KI und/als bildende Kunst. In *Handbuch Künstliche Intelligenz und die Künste*, Hg. Stephanie Catani und Jasmin Pfeiffer. Berlin: De Gruyter.

Radford, Alec, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry et al. 2021. Learning Transferable Visual Models from Natural Language Supervision. https://doi.org/10.48550/arXiv.2103.00020.

Radford, Alec, Luke Metz und Soumith Chintala. 2015. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. https://doi.org/10.48550/arXiv.1511.06434.

Rombach, Robin, Andreas Blattmann, Dominik Lorenz, Patrick Esser und Björn Ommer. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. https://doi.org/10.48550/arXiv.2112.10752.

Underwood, Ted. 2021. Science Fiction Hasn't Prepared Us to Imagine Machine Learning. Blog: *The Stone and the Shell*. https://tedunderwood.com/2021/02/02/why-sf-hasnt-prepared-us-to-imagine-machine-learning/. Zugegriffen: 15. November 2022.