



Hochschule Macromedia für angewandte
Wissenschaften,
University of Applied Sciences

BACHELORARBEIT

zur Erlangung des akademischen Grades
Bachelor of Arts

Künstliche Intelligenz
im Kontext der 3D Environmentgestaltung -
Integrationsmöglichkeiten und Automatisierung
- Stand 2023/24

im Studiengang Design
Studienrichtung Animation

Erstprüfer:
Prof. Karin Jobst

Vorgelegt von:

Jonas Lederer

Matr.-Nr.: M-43394

Studiengang: Design (B.A.)

Fachrichtung: Animation

München, im Januar, 2024

Inhaltsverzeichnis

Abstract

- 1.0 Einleitung..... 1
 - 1.1 Relevanz des Themas 1
 - 1.2 Fragestellung und Zielsetzung 1
 - 1.3 Methodik 2
- 2.0 Theoretische Grundlage 3
 - 2.1 Virtuelle Welten und 3D Environments..... 3
 - 2.2 Künstliche Intelligenz 5
 - 2.2.1 Large Language Models 6
 - 2.2.2 Diffusion Modelle und Diffuser Netzwerke 6
 - 2.2.3 ControlNet..... 7
 - 2.2.4 Neural Radiance Fields - NeRF 7
 - 2.2.5 CLIP 8
- 3.0 Ergebnisse 8
 - 3.1 Moodboards, Referenzfotos, Styleframing 8
 - 3.2 Hintergründe und Beleuchtung..... 9
 - 3.3 Camera Projection Method 10
 - 3.4 Texturieren mit AI11
 - 3.5 3D Modeling mit AI..... 12
 - 3.7 KI Scene Generation..... 16
 - 3.8 Large Language Models 18
- 4.0 Diskussion 18
 - 4.1 Integration von Künstlicher Intelligenz in den Arbeitsprozess 18
 - 4.2 Ablösung des 3D Environment Artists durch Künstliche Intelligenz..... 21
- 5.0 Fazit 22

- Literaturverzeichnis 24
- Abbildungsverzeichnis..... 28
- KI Verzeichnis..... 29
- Eidesstattliche Erklärung..... 30

Abstract

Diese Arbeit beschäftigt sich mit der Integration von KI im Kontext der 3D Environment Gestaltung. Sie untersucht wie KI im Zeitraum 2023-2024 in den Arbeitsprozess von 3D Environment Artists integriert werden kann und inwieweit diese durch KI ersetzt werden könnten. Zu diesem Zweck kombiniert die Arbeit eine umfassende Literaturrecherche mit Experteninterviews und Praktischen Experimenten, um das Potential von KI Technologie in diesem Bereich zu bewerten. Es werden verschiedene KI Anwendungen betrachtet, die im Kontext der 3D Environment Gestaltung aktuell relevant sind. Dabei wurden verschiedene Aufgabengebiete der 3D Environment Gestaltung identifiziert in denen diese Anwendungen verwendet werden können.

1.0 Einleitung

1.1 Relevanz des Themas

Künstliche Intelligenz (KI) ist das große Thema der Jahrzehntwende. Seit der Veröffentlichung von ChatGPT im Juni 2020 ist die öffentliche Aufmerksamkeit stark auf das Thema KI gefallen. Durch das Debüt von Bildgeneratoren, wie dem Text-to-Image-Modell von Midjourney im Juni 2022, das mittlerweile über 16 Millionen registrierte Nutzer hat (Dawood, 2023), wird das Potential, das sich hinter dem Begriff „Künstliche Intelligenz“ verbirgt, visuell greifbar.

Seit den unglaublichen Erfolgen von OpenAI, dem Entwickler von ChatGPT, und Midjourney, nutzen viele Entwickler KI, um neue Türen in verschiedene Bereiche zu eröffnen, so auch in der 3D Computer-Grafik. Nvidia, einer der populärsten Grafikkarten- und Softwarehersteller, arbeitet beispielsweise parallel an mehreren KI-Technologien und beschäftigt ein eigenes Team an Wissenschaftlern, das sich speziell mit der Erforschung von KI in der Computer Grafik beschäftigt (NVIDIA Corporation, o. D.-B). Aber nicht nur große Entwicklerstudios wie Nvidia arbeiten an Software, die 3D Artists bei ihrer Arbeit unterstützen könnten. Auch auf den Marktplätzen etablierter 3D Software, wie dem Blender Market (Blender Market, o. D.) oder dem Marktplatz der Unreal Engine (Unreal Engine Marketplace, o. D.), finden sich immer mehr Add-ons, die KI nutzen, um 3D-Artists bei ihrer Arbeit zu unterstützen.

1.2 Fragestellung und Zielsetzung

Durch die Aktualität des Themas und die rasanten Entwicklungen im Bereich der Anwendung von KI-Technologien machen sich viele Designer Sorgen um die Zukunft ihres Arbeitsplatzes (Neuburg et al., 2020), so auch 3D-Artists. Aussagen aus Berichten wie dem von Hatzius et al. (2023), nach denen rund ein Viertel der derzeitigen Arbeitsplätze weltweit durch KI ersetzt werden könnten, drängen zu einer Auseinandersetzung mit der Thematik des Einsatzes von KI im Arbeitsumfeld. Auf Grund der schöpferischen Natur von generativen KIs wie Dall-E, Midjourney oder Stable Diffusion ist es besonders für Designer von großer Wichtigkeit zu wissen, wie KI ihre Arbeitsweise beeinflussen kann. Es gibt bereits erste Literaturarbeiten zu generativer KI in 3D (Li et al., 2023), jedoch haben sich bis jetzt noch keine Autoren mit dem aktuellen Forschungsstand von KI im Kontext spezieller Bereiche innerhalb des 3D-Designs beschäftigt. Die Forschungsfrage dieser Arbeit lautet deshalb:

„Wie wird Künstliche Intelligenz aktuell in den Arbeitsprozess von 3D-Environment-Artists integriert und inwieweit könnten diese durch Künstliche Intelligenz ersetzt werden?“

Ziel dieser Bachelorarbeit ist es, das Potential von künstlicher Intelligenz im Kontext der 3D Environment Gestaltung zu überprüfen. Es sollen konkrete Aufgaben identifiziert werden, die Stand 2023 - 2024 von KI übernommen werden können oder welche, die durch die Verwendung von KI beeinflusst werden, wie z.B. das Erstellen von Texturen oder Modellen.

1.3 Methodik

Im Zuge dieser Arbeit wurde eine themenumfassende Literaturrecherche durchgeführt. Diese dient als Grundlage, um umfassendes theoretisches Wissen über das Thema zu sammeln und so einen Überblick über die im Zeitraum 2023 - 2024 verfügbaren KI-Technologien, die 3D Artists bei der Gestaltung von virtuellen Umgebungen verwenden können, zu schaffen. Dafür wurden die Datenbanken Researchgate, Springerlink, arxiv, IEEE Xplore, sowie Google Scholar nach den folgenden Stichwörtern durchsucht:

Generative AI, 3D, CGI, 3D-Modeling, Texturing, Worldbuilding, Virtual Worlds und 3D-Environment-Design

Zudem wurden Publikationen bekannter Softwareunternehmen, die sich mit der Erforschung von KI beschäftigen, miteinbezogen. Nach einem Screening von Abstracts und Fazits der Papers wurden dann relevante Paper identifiziert. Die ausgewählten Papers wurden auf die Anwendbarkeit in der 3D Environment Gestaltung untersucht und die Ergebnisse systematisch zusammengefasst. Da die genauen technischen Hintergründe und Abläufe von KI jedoch sehr komplex und nicht zielführend für diese Arbeit sind, konzentriert sich die Analyse der Papers vor allem auf die praktische Anwendbarkeit von KI zur 3D Environment Gestaltung.

Darüber hinaus wurden zwei Experteninterviews durchgeführt, die praxisorientierte Informationen über Integrationsmöglichkeiten und technische Ansätze liefern sollten. Die Experteninterviews wurden anhand eines zuvor erstellten Leitfadens durchgeführt, der genau wie die Transkriptionen der Interviews, den Zusatzmaterialien zu entnehmen ist. Die Fragen fokussierten sich neben den praktischen Anwendungsmöglichkeiten von KI auch auf die durch KI ausgelöste Entwicklung in der 3D Branche.

Das erste Interview wurde mit einem freiberuflichen CG-Artist mit über 15 Jahren Berufserfahrung durchgeführt. In der Arbeit wird er als „Interviewpartner Eins“ zitiert. Nach einem Studium an der Bayrischen Filmakademie war er lange Zeit bei einer Münchner Medienproduktion als Artdirector für 3D Animation tätig. Er beschreibt sich selbst als 3D-Generalist, und hat weitreichende Erfahrung in Animation, Visualisierung, Visual Effects und Compositing. Im Zuge seiner freiberuflichen Tätigkeit organisiert er außerdem Veranstaltungen, bei denen sich Künstler und Designer treffen, um über ihren Einsatz von KI im Arbeitsalltag zu berichten.

Der Partner des zweiten Interviews, der in der Arbeit als „Interviewpartner Zwei“ zitiert wird, beschreibt sich selbst als CG-Generalist mit Fokus auf Technical Art, Prozeduralismus und Automatisierung. Nach einem abgeschlossenen Design Studium an einer der renommiertesten Universitäten Deutschlands, war er neben seiner freiberuflichen Tätigkeit im Bereich Motion Design und Procedural Modeling an mehreren Hochschulen und Universitäten als Dozent tätig. Mit KI beschäftigt er sich seit 2016.

Die Arbeit wird zudem durch das praktische Erproben der recherchierten Tools und somit der eigenen Expertise des Autors ergänzt. Dafür wurden die recherchierten Technologien in der Vorbereitung zu Praxisarbeit erprobt und dokumentiert. Genauere Informationen sind der Dokumentation der praktischen Arbeit zu entnehmen.

Die Kombination aus Fachliteratur, experimenteller Forschung und Expertenmeinungen gewährleistet, dass die Ergebnisse der Arbeit auf einer soliden theoretischen Basis beruhen. Sie bietet ein breit gefächertes Spektrum an Informationen, die sich auf Theorie und Praxis bezieht und bietet somit eine Momentaufnahme der sich rasant weiterentwickelnden KI-Technologien.

2.0 Theoretische Grundlage

2.1 Virtuelle Welten und 3D Environments

In seinem Buch „Designing Virtual Worlds“ definiert der Autor und Videospiel-Forscher Richard A. Bartel (2003) virtuelle Welten als „places where the imaginary meets the real“ (S.1). Um diese abstrakte Definition besser zu erklären, gibt Bartle fünf Konventionen an, denen virtuelle Welten unterliegen, die sie von den verwandten, nichtrealen Räumen unterscheiden (Bartle, 2003, S.4-5):

- Der Welt liegen automatisierte Regeln zugrunde, die es den Spielern ermöglichen, Änderungen an der Welt, nicht aber an den Regeln selbst, vorzunehmen. Er beschreibt dies als „the Worlds Physics“.
- Spieler repräsentieren Individuen innerhalb der Welt, mit denen sie sich sehr stark identifizieren. Den sogenannten Charakter. Dieser Charakter kann entweder teilweise oder vollständigen Einfluss auf die Virtuelle Welt haben. Die gesamte Interaktion innerhalb der Virtuellen Welt wird über Charaktere kanalisiert.
- Interaktion mit der Welt findet in Echtzeit statt. Wenn man etwas in der Welt macht, kann man sofortiges Feedback erwarten.
- Die Welt wird geteilt.
- Die Welt ist zumindest zu einem gewissen Grad hartnäckig.

Eine andere Definition virtueller Welten von Schroeder (2008) lautet „Virtual worlds are persistent virtual environments in which people experience others as being there with them - and where they can interact with them“ (Schroeder, 2008, S.2). Der Begriff virtuelle Welt meint also eine virtuelle Umgebung im digitalen Raum, in der Menschen mit anderen Menschen und/oder der Umgebung interagieren können.

Doch wie schaut die virtuelle Umgebung einer solchen virtuellen Welt eigentlich aus? Die Umgebungen der ersten virtuellen Welten waren textbasiert, was bedeutet, dass diese nicht mit Bildern sondern mit Worten dargestellt wurden (Bartle, 2003, S.3). Ein abstrakter Gedanke, wenn man bedenkt, dass moderne Computer durch die dreidimensionale Darstellung von hochpolygonalen Geometrien, hochauflösenden Texturen und durch moderne Rendersoftware dazu in der Lage sind, Bilder zu errechnen, die kaum noch von der Realität zu unterscheiden sind.

Mit den Entwicklungen in der Computerindustrie haben sich also auch die Möglichkeiten weiterentwickelt, virtuelle Umgebungen zu gestalten. Für die Gestaltung von Umgebungen innerhalb virtueller Welten, wenden sich die Entwickler heutzutage in vielen Fällen an 3D-Artists, die sich speziell mit der Gestaltung der Umgebung auseinandersetzen, den 3D-Environment-Artists.

Der Begriff 3D Environment ist dabei folgendermaßen definiert: „A 3D Environment is a three-dimensional representation of geometric data that resembles the real world. It is simply a digital setting or background created with computer graphics software. It can be an indoor or outdoor location with realistic or non realistic objects that give you visual sensations“ (ProGlobalBusinessSolutions, 2023). Der Begriff 3D Environment bezieht sich demnach also auf die dreidimensionale Darstellung von Räumen oder Szenarien in der Computergrafik. Es handelt sich also um die grafische Gestaltung einer digitalen Kulisse, wie zum Beispiel der Umgebung in einem Computerspiel oder virtuellen Welten.

Die Geometrien, die dabei verwendet werden, werden in der Regel über polygonale Netzstrukturen visualisiert, die 3D Meshes genannt werden. Diese Meshes setzen sich aus vielen Eckpunkten (Vertices), Kanten (Edges) und Flächen (Polygone) zusammen. Zwei miteinander verbundene Eckpunkte ergeben eine Kante. Verbindet man drei oder mehr dieser Eckpunkte entsteht zwischen diesen eine Fläche, die man Polygon nennt (Danthree Studio, 2024). Neben der mesh-basierten Darstellung von Objekten können Geometrien auch über Point Clouds, eine Ansammlung von Millionen einzelner Punkte, die jeweils eine x-, y- und z-Koordinate besitzen und in Kombination eine Geometrie erzeugen, dargestellt werden (Higgins, 2021).

Dreidimensionale Umgebungen finden jedoch nicht nur in Computerspielen und virtuellen Welten Anwendung, sondern sind auch in der Filmproduktion, Produktpräsentation

oder in der Visualisierung von Architektur vorzufinden. Wie ein Environment gestaltet wird, hängt stark davon ab für welchen Zweck es erstellt wird.

Die Umgebungen immersiver Erfahrungen, wie bei frei begehbaren Open-World-Spielen, sind besonders aufwendig zu produzieren, da sich die Nutzer frei bewegen können und so auch selbst die Perspektive bestimmen, aus der sie auf diese blicken. Die Designer müssen also die gesamte Welt und alle sich in ihr befindlichen Objekte dreidimensional aufbauen, um zu gewährleisten, dass sie aus jeder eingenommenen Perspektive funktioniert.

Anders sieht es bei Umgebungen aus, die in einem Kontext entworfen werden, in dem ein Kamerawinkel definiert ist, wie zum Beispiel in Animationsfilmen oder in manchen Videospielen. Da der Blickwinkel des Betrachters auf die dargestellte Umgebung fest definiert ist, müssen Geometrien innerhalb des dargestellten Bildes nur aus dieser einen Perspektive funktionieren, wodurch sich in der Gestaltung neue Möglichkeiten eröffnen, wie z.B. die Verwendung von zweidimensionalen Daten in einem dreidimensionalen Umfeld (Mickmumpitz, 2023).

2.2 Künstliche Intelligenz

Auf der Webseite des europäischen Parlaments wird Künstliche Intelligenz als die Fähigkeit einer Maschine beschrieben, menschliche Fähigkeiten wie logisches Denken, Lernen, Planen und Kreativität zu imitieren (*Was ist Künstliche Intelligenz und wie wird sie genutzt?*, 2020). Schaut man sich aber wissenschaftliche Publikationen zu diesem Thema an, fällt auf, dass es keine einheitliche Definition für den Begriff gibt.

Im Zusammenhang mit KI fallen häufig die Begrifflichkeiten wie Deep- oder Machine Learning. Diese beschreiben alle artverwandten Technologien, sind aber nicht synonym, wie Phil Wennker (2020, S.9) in seinem Buch „Künstliche Intelligenz in der Praxis“ erklärt. Weiterhin schreibt er „Während AI oder künstliche Intelligenz eine Art Mantelbegriff ist, der alle Anwendungen zusammenfasst, die in irgendeiner Form Intelligenz zeigen, ist Machine Learning der Prozess aus Daten zu lernen und Schlüsse zu ziehen. Deep Learning ist ein Teilbereich des Machine Learning der auf tiefe künstliche neuronale Netze setzt, um kontinuierlich zu lernen und Erlerntes bei weiteren Interaktionen mit der Welt anzupassen und zu modifizieren“ (Wennker, 2020, S.9).

Neuronale Netze müssen mit Datensätzen trainiert werden (Raveling, 2022), um einen gewünschten Output zu erzeugen und decken derzeit beinahe alle für 3D-Artists relevanten Methoden und Werkzeuge ab (Interviewpartner Zwei, 2024, S.1). Wie genau diese neuronalen Netze funktionieren ist jedoch zu komplex und für diese Arbeit nicht relevant, weswegen an dieser Stelle auf eine genauere Erklärung der Funktionsweise verzichtet wird.

Im folgenden Teil der Arbeit werden stattdessen Kern-Technologien, sowie wichtige Begrifflichkeiten aus aktuell relevanten Publikationen im Bereich KI in der Computergrafik gesammelt und erklärt. Wie diese Kerntechnologien weiterentwickelt, miteinander verknüpft und von 3D-Environment-Artists genutzt werden können, wird später im Ergebnis teil der Arbeit festgehalten. Angesichts der rasanten Technologieentwicklung in diesem Bereich handelt es sich bei dieser Arbeit um eine Momentaufnahme des Zeitraums zwischen Oktober 2023 und Januar 2024.

2.2.1 Large Language Models

Large Language Models (LLMs) basieren auf einer neuronalen Netzwerkarchitektur, die Transformer genannt wird (Muehmel, 2023), weswegen LLMs auch häufig Transformer Netzwerke genannt werden. Zu den LLMs gehören alle state-of-the-art Textgeneratoren, also alle GPT ähnlichen Modelle (Interviewpartner Zwei, 2024). „LLMs beschränken sich bei Input und Output auf Text, können innerhalb dieses Formats aber sehr viele Stile abbilden, wie Konversationen, Programmcode in diversen Sprachen, Übersetzungen und diverse literarische Stile“ (Interviewpartner Zwei, 2024, S.1).

2.2.2 Diffusion Modelle und Diffuser Netzwerke

Diffusion Modelle, auch bekannt als „Denoising diffusion probabilistic models“ (DDPMs) oder „score based generative models“, sind eine Klasse der generativen KI-Modelle, die dazu in der Lage sind, hochauflösende Bilder zu generieren (Ahirwar, 2023). Diese Bilder, teilweise auch Videos, werden aus kurzen beschreibenden Texten, die Prompts genannt werden, generiert (Oppenlaender, 2023). Die neu generierten Daten ähneln dabei denen, die verwendet wurden, um sie zu trainieren (Li et al., 2023, S.5), was dazu führt, dass viele Diffusionsmodelle, wie beispielsweise Stable Diffusion XL (SDXL), in selbige sind. „SDXL kann mir perfekte Portraitfotos erzeugen, scheitert aber noch immer bei Händen. Es kann mir wunderschöne Fantasy-Landschaften generieren, weiß aber nicht, welche Komponenten für ein Rennrad gezeichnet werden müssen“ (Interviewpartner Zwei, 2024, S.3).

Diffusionsmodelle durchlaufen beim Generieren neuer Bilder zwei Schritte: Den „Forward diffusion process“, wobei dem Originalbild in kleinen Schritten ein Rauschen hinzugefügt wird, und den „Reverse diffusion process“, bei dem ein neuronales Netz darauf trainiert wird, die ursprüngliche Datei wieder herzustellen, indem es den im „Forward diffusion process“ angewendeten Rauschprozess umkehrt (Ahirwar, 2023).

Diffuser Netzwerke lassen sich zusammenfassend als „Input-Output-Systeme: Text-to-Image, Image-to-Image, Text-to-Video, Image-to-Video, Text-to-3D Model, etc.“ (Interviewpartner Zwei, 2024, S.1) beschreiben.

2.2.3 ControlNet

ControlNet ist eine „neural network architecture to add spatial conditioning controls to large, pretrained text-to-image diffusion models“ (Zhang & Agrawala, 2023). Die Technologie bietet also eine Möglichkeit Diffusion Modelle besser zu kontrollieren, indem „extra conditions“ ergänzt werden. Konkret bedeutet das, dass Bilddaten hochgeladen werden können, an denen sich Stable Diffusion orientiert. So können beispielsweise Komposition, Stil und Farbe des ausgegebenen Bilds beeinflusst werden, wodurch viel präzisere Ergebnisse erzielt werden können.

2.2.4 Neural Radiance Fields - NeRF

Die Entwicklungen von KI in den letzten Jahren haben auch Möglichkeiten ergeben, wie KI dazu verwendet werden kann, um Objekte im digitalen dreidimensionalen Raum darzustellen. Eine immer bekannter werdende Technologie in diesem Bereich heißt NeRF. NeRF ist ein Akronym für Neural Radiance Fields. In dem 2020 erschienen Paper mit dem Namen „NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis“ präsentieren Mildenhall et al. (2020) eine neue Möglichkeit komplexe Szenen im digitalen dreidimensionalen Raum nachzubilden. Hierzu werden Fotos von einem Objekt oder einem Raum aus verschiedenen Perspektiven aufgenommen, die dann dazu verwendet werden, eine virtuelle volumetrische Repräsentation der aufgenommenen Szene zu erstellen (siehe Abb. 1). Diese Technik erinnert im ersten Moment sehr an die schon länger existierende Technologie der Fotogrammetrie. Im Gegensatz zur Fotogrammetrie verwendet die NeRF-Technologie jedoch ein neuronales Netz, das aus den für die Szene aufgenommenen Bildern lernt, wie das dargestellte Objekt aussehen würde, wenn man es aus einer Perspektive betrachtet, die man zuvor nicht fotografiert hat. Besonders ist dabei, dass die NeRF-Technologie nicht nur einfach ein 3D Modell erstellt, sondern Oberflächeneigenschaften wie Reflexionen oder Farben berücksichtigt, die sich mit dem Blickwinkel verändern. Dadurch können sehr realistische Ergebnisse erzielt werden. Eine Besonderheit dieser Technologie ist hierbei, dass die KI bereits aus wenigen Eingabebildern vorzeigbare Ergebnisse liefert.



Abbildung 1: Neural Radiance Field (Mildenhall et al. 2020, S.2)

2.2.5 CLIP

CLIP (Contrastive Language-Image Pre-training) ist eine von OpenAI entwickelte KI-Technologie, die erstmals in „Learning Transferable Visual Models From Natural Language Supervision“ von Radford et al. (2021) vorgestellt worden ist. Vereinfacht gesagt ist Clip ein neuronales Netz, das dazu in der Lage ist, Bilder zu beschreiben. Besonders hierbei ist, dass CLIP auch in der Lage ist Bilder zu beschreiben, die sich stark von den Datensätzen unterscheiden, die verwendet wurden, um das Modell zu trainieren (Brems, o. D.).

3.0 Ergebnisse

3.1 Moodboards, Referenzfotos, Styleframing

Eine der wohl am naheliegendsten Verwendungszwecke von KI in den meisten Bereichen des Designs und so auch in der Gestaltung von 3D Environments, ist die Erstellung von Moodboards, Layouts und allgemein Referenzfotos. Hierfür können Artists generative Bildmodelle wie Midjourney, Dall-E oder Stable Diffusion verwenden. Das Besondere beim Verwenden dieser Modelle ist, dass Referenzfotos über Text und Bildeingabe an spezielle Bedürfnisse angepasst werden können. So kann im Vorfeld der Produktion bereits ein Bild des Endproduktes aufgezeigt werden. „Im Prinzip ist es ein bisschen so wie Googlen, aber man kann es halt dann auch schön spezifizieren“ (Interviewpartner Eins, 2023, Min.00:05.01).

Eine Applikation, die für diesen Zweck entwickelt worden ist, heißt 3Dall-E. 3Dall-E ist ein Plugin entwickelt für die 3D CAD Software Fusion 360, das Text-to-Image KI in den CAD und Produktdesign Workflow integriert (Vivian Liu et al., 2023). Das Plugin kombiniert Dall-E, GPT-3 sowie CLIP und ermöglicht den Designern Bilder zu generieren, die ihnen als Inspirationsquelle für den Arbeitsprozess dienen. Neben der Bildgeneration basierend auf Text („Text-to-Image“) beinhaltet das Plugin auch eine „image-prompt“ Funktion, die auf einem Rendering des aktuellen Workspace basiert und so eine Inspirationsquelle für weitere Arbeitsschritte darstellt.

Auch Interviewpartner Zwei (2024) hat angegeben, dass diffusion-basierte Bild- und Videogeneratoren aktuell besonders in der Styleframing Phase eingesetzt werden. „Den Haupt-Einsatzzweck von Diffuser-Modellen sehe ich momentan in der Styleframing Phase von Projekten. Das Ziel dieser Phase ist, dem Kunden über Bilder und kurze Videos eine Vorstellung vom fertigen Film zu geben“ (Interviewpartner Zwei, 2024, S.3). Auch wenn hier von einer Filmproduktion gesprochen wird, ist die Styleframing Phase ein wichtiger Teil vieler Produktionen, in denen 3D-Umgebungen Teil des Endproduktes sind (z.B. in der Videospiel Produktion). Für 3D-Artists bedeutet das Arbeiten in der

Styleframing Phase das schnelle und meist schlampige Erstellen von vielen Renderings, eine Arbeit, die nicht besondere kreativ ist, da es meist einfach nur darum geht, einen bestehenden Stil auf den Film oder das Spiel zu übertragen (Interviewpartner Zwei, 2024, S.3). „Aus all diesen Gründen können Diffuser-Modelle hier jetzt schon eingesetzt werden und werden in naher Zukunft mit besseren Video-Modellen nur noch mehr Arbeit abnehmen“ (Interviewpartner Zwei, 2024, S.3).

3.2 Hintergründe und Beleuchtung

Hintergründe sind ein wichtiges Thema und ein breit gefächelter Begriff, wenn man sich mit der 3D Environment Gestaltung auseinandersetzt. In Filmen dienen die von den Artists gestalteten Umgebungen häufig als Hintergrund einer sich im Vordergrund abspielenden Scene, die vor einem Greenscreen gedreht worden ist (siehe Abb. 2). Das 3D Environment wird auf der Scene basierend aufgebaut und in der Postproduktion des Filmes als Hintergrund eingefügt.



Abbildung 2: 3D Environment als Hintergrund (Digital Synopsis, 2023)

Aber auch innerhalb der 3D Environments selbst, unabhängig vom Produktionskontext, werden häufig Hintergründe verwendet. Das hat vor allem den Grund, dass man die Umgebungen nicht bis zum Horizont aufbauen kann. Hintergründe werden dann meist in Form von 360 Grad Fotografien über einen Sky-Dome (auch Sky-Box genannt), eine Art riesige Kugel, die die gesamte Szene umschließt, in die Umgebung integriert (siehe Abb. 3). Durch die Verwendung von Skyboxes kann die Illusion einer weitläufigen Landschaft geschaffen werden ohne detaillierte 3D Geometrien für weit entfernte Objekte, wie Berge oder Wolken, zu verwenden (Blockade Labs, o. D.).



Abbildung 3: Sky-Dome (in Anlehnung an Blockade Labs, o. D.)

Ein besonderer Vorteil bei der Verwendung dieser Technik ist, dass die verwendeten Bilddateien (im Normalfall im HDRI-Format) gleichzeitig als Lichtquelle dienen und so eine realistische Beleuchtung der gesamten Szene ermöglichen. Die Beleuchtung ist ein ausschlaggebender Faktor in der Gestaltung eines 3D Environments, da das Licht einen großen Einfluss auf die Atmosphäre der Umgebung hat.

Die Recherchen haben zwei aktuell relevante Tools ergeben, die eine auf KI basierende Produktion dieser 360 Grad Bilder ermöglicht: „Skybox AI“ (Blockade Labs, o. D.) und „NVIDIA Canvas“ (NVIDIA Corporation, o. D.-A).

Skybox AI ist eine webbasierte Applikation der Firma „Blockade Labs“, die die Möglichkeit bietet 360 Grad Bilder durch die Eingabe von Textprompts zu erstellen (siehe Abb. 4). Dabei verwendet Skybox AI laut eigener Aussage, eine modifizierte Version von Stable Diffusion (Blockade Labs, o. D.). In der Applikation kann vor der Generation des 360 Grad Fotos zwischen einer Vielzahl an Stilen, wie Realistic, Fantasy oder Futuristic gewählt werden. Außerdem kann man über einen Painting Modus gewünschte Objekte einzeichnen, die in der Generation berücksichtigt werden.



Abbildung 4: 360 Grad Foto generiert durch Skybox AI (Blockade Labs, o.D)

3.3 Camera Projection Method

Die Recherchen zu weiteren Anwendungen von Diffusion Modellen, führten zu einem Video, in dem ein alternativer Ansatz zur Gestaltung von 3D Umgebungen präsentiert wird (Mickmumpitz, 2023). Die in dem Video vorgestellte Methode präsentiert einen Workflow zwischen der 3D Software Blender, Stable Diffusion und ControlNet. Der Artist beginnt bei der Erstellung des Environments, wie in den meisten herkömmlichen Workflows, mit einem groben Blockout der Umgebung und einer Kamera, die die Szene einfängt. Von dieser Kamara ausgehend wird eine z-depth-map gerendert (siehe Abb. 5).

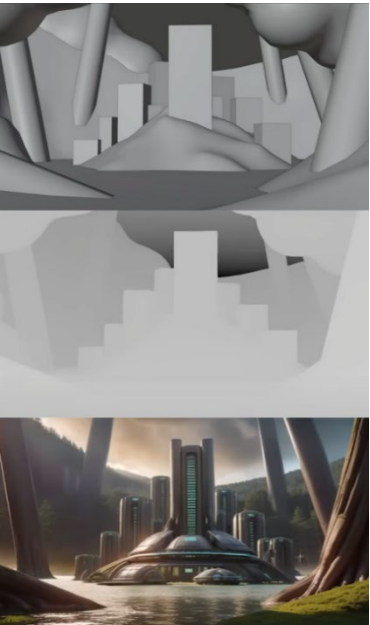


Abbildung 5: Blockout, Depthmap & AI Bild (Mickmumpitz, 2023)

Nach der Definition von Goulekas (2001) ist eine depth-map „A greyscale image in which the pixels represents the distance from camera where each object in a 3D scene resides, much the way a alpha channel stores information regarding the opacity for each pixel“, also eine schwarz-weiße Bilddatei, die die Position von 3D Geometrie im dreidimensionalen Raum bestimmt.

ControlNet ist dazu im Stande diese Bilddatei zu verstehen, weswegen diese in Stable Diffusion verwendet werden kann, um präzise Befehle für die Gestaltung eines Bildes zu geben. Das vom Artist gepromptete Bild wird dann in Blender importiert und über eine Funktion namens „Camera Projection“ auf die zu Beginn geblockte Geometrie projiziert. So entsteht eine dreidimensionale Repräsentation des durch die KI generierten Bildes, in der die Kamera und Geometrien zumindest zu einem gewissen Grad bewegt und neu angeordnet werden können.

Da bei dieser Methode eine zweidimensionale Bilddatei aus der Perspektive der Kamera auf die Geometrie übertragen wird, ist es jedoch nicht möglich gänzlich neue Perspektiven einzunehmen. Eine freie Bewegung im Raum, wie zum Beispiel eine 360 Grad Kamerafahrt um ein Objekt im Fokus, ist also nicht möglich, da die Bilddatei als Textur aus dem zu Beginn definierten Kamerawinkel auf die Objekte der Szene projiziert worden ist. Zu starke Bewegungen der Kamera resultieren also in einer Verzerrung der Texturen (siehe Abb. 6).

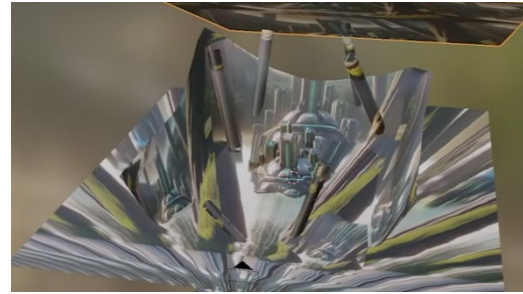


Abbildung 6: Verzerrte Texturen (Mickmumpitz, 2023, Min. 08:08)

3.4 Texturieren mit AI

Die Texturierung der 3D Geometrien, aus denen ein Environment aufgebaut ist hat einen großen Einfluss auf die dem Betrachter kommunizierte Atmosphäre und Stimmung und gehört zu den zentralen Aufgaben von 3D Environment Artists. In der Texturierung werden den Objekten Farbigkeit, Musterung und Oberflächenbeschaffenheit wie z.B. Reflexionsgrad hinzugefügt (*Was ist 3D-Texturierung?* | Adobe Substance 3D, o. D.). In 3D Programmen werden diese Parameter meist über 2D Bilddaten auf das 3D Objekt projiziert. Hierzu wird meist eine UV Map, eine zweidimensionale Repräsentation der 3D Geometrie, verwendet.

In der Wissenschaft lassen sich zahllose Publikationen finden, die Methoden vorstellen, in denen KI verwendet wird, um Artists bei der Texturierung ihrer 3D Modelle zu unterstützen (Richardson et al., 2023; Chen et al., 2022; R. Chen et al., 2023; D. Z. Chen et al., 2023). Bei Betrachtung dieser Publikation fällt auf, dass alle Modelle auf Texteingabe bzw. der Eingabe von Text Prompts basieren. Jedoch unterscheiden sich die vorgestellten Methoden in ihren spezifischen Anwendungsbereichen und technischen Ansätzen. TANGO fokussiert sich beispielsweise „on stylizing photorealistic appearance renderings of a given surface mesh of arbitrary topolog“ (Y. Chen et al., 2022), also auf die Synthese realitätsnaher Texturen. Text2Tex hingegen bietet einen neuen technischen Ansatz zur Generierung von Texturen, die Ansammlungen von inkonsistenten und gestreckten Artefakten auf den Texturen vermeiden soll (D. Z. Chen et al., 2023).

Um die in den Publikationen präsentierten Frameworks verwenden zu können, sind jedoch teilweise hohe Systemanforderungen sowie ein grundlegendes Verständnis von Programmiersprache vorausgesetzt, wodurch sich die Anwendung der vorgestellten Frameworks für viele Artists schwierig gestaltet. Die Recherchen haben allerdings auch Möglichkeiten ergeben, welche diese oder ähnliche Technologien einfach zugänglich machen. Eine dieser Möglichkeiten ist das 3D Texture Generation Tool von Leonardo.Ai

(Leonardo Interactive Pty Ltd, o. D.). In dieser browserbasierten Applikation können 3D Modelle hochgeladen und über Texteingabe vollautomatisch texturiert werden. Eine weitere nutzerfreundliche KI-Lösung zur Generation von Texturen ist das von dem Github-nutzer „carson-katri“ veröffentlichte Blender Plugin namens „dream-textures“, dessen letzte Version „v.0.3.1“ am 7. November 2023 veröffentlicht worden ist (Carson-Katri, 2023). Dieses Plugin ermöglicht es, innerhalb von Blender Diffusionsmodelle herunterzuladen und diese zum Generieren von Texturen zu verwenden. Auch hier werden wieder Textprompts als Eingabemethode verwendet, es gibt jedoch auch die Möglichkeit „seamless textures“ zu generieren, eine Voraussetzung die Texturen in vielen Anwendungsfällen erfüllen müssen.

„Seamless“ oder „tilable“ Texturen sind Bilddateien, die neben sich selbst platziert werden können (oben, unten, rechts, links) ohne einen offensichtlichen „seam“, also eine Naht zu hinterlassen. Das ist besonders wichtig, wenn Texturen skaliert, also an die Größe des 3D Objektes angepasst werden müssen.

Viele der aktuell etablierten Text-zu-Bild Generatoren scheitern beim Versuch solche seamless Texturen zu erstellen. Midjourney hat allerdings eine Funktion, die das Erstellen von tilable Images ermöglicht. Möchte man also mit Midjourney eine seamless Textur erstellen, kann man in Midjourney den Parameter „--tile“ an den Textprompt anhängen und das generierte Bild wird seamless (*Midjourney tile parameter*, o. D.)(siehe Abb. 7).

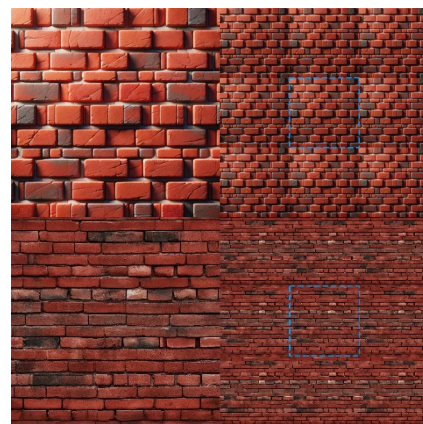


Abbildung 7: Seamless und nicht seamless Texturen generiert mit Dall E (oben) und Midjourney (unten) (eigene Darstellung mit Elementen generiert durch Midjourney und Dall-E)

3.5 3D Modeling mit AI

3D Environments werden aus vielen verschiedenen 3D Modellen zusammengesetzt, die in Kombination das Environment ergeben. Mit den technischen Entwicklungen wurden im Laufe der Zeit verschiedene Methoden entwickelt, um solche 3D Modelle zu erstellen. Neben „traditionellen“ 3D Modellierungstechniken wie dem Sculpting, wurde auch eine Methode entwickelt, um aus zweidimensionalen Fotos 3D-Daten zu extrahieren: Die Fotogrammetrie (*Photogrammetrie-Software – Fotos in 3D-Scans | Autodesk*, o. D.). Durch die kürzlichen Entwicklungen im Bereich des Machine Learnings und den daraus entstandenen Modellen zur Generierung von 2D Bildern, wurden auch auf KI basierende Methoden zu Erstellung von 3D Modellen entwickelt. „Due to the development in text-to-image as well 3D modeling technologies (like NeRF), text-to-3D has become a newly

emerging yet highly active research field.“ (2023). Während der Recherche dieser Arbeit wurden eine Vielzahl an Publikationen gefunden, die solche Methoden vorstellen (Poole et al., 2022; Lin et al., 2023; Nichol et al., 2022; Saito et al., 2020; J. Gao et al., 2022). Die aufgeführten Publikationen präsentierten Frameworks, die sich in zwei Kategorien aufgliedern lassen: Text-to-3D Modelle und Image-to-3D Modelle.

Text-to-3D Modelle

Zu den Text-to-3D Modellen gehören Modelle, wie das von Poole et al. vorgestellte „DreamFusion“ (2022), „Magic3D“ (Lin et al., 2023) und „Point-E“ (Nichol et al., 2022). Diese Modelle stellen alle Methoden zur Synthese von texturierten 3D Modellen durch Texteingabe dar. Um diese Synthese durchzuführen, werden in den Publikationen unterschiedliche Ansätze präsentiert.

Um 3D Modelle über Diffusion Modelle direkt erstellen zu können, wären enorme Datensätze von gelabelten 3D Modellen nötig, welche es aktuell noch nicht gibt. Poole et al. (2022, S. 1) umgehen diese Problematik, indem sie ein vorab trainiertes 2D Text-zu-Bild Diffusionsmodell verwenden, um die Text-zu-3D Synthese durchzuführen. Die mit „DreamFusion“ (Poole et al., 2022) präsentierte Methode kombiniert dabei zwei der in dieser Arbeit bereits behandelten KI-Modelle (siehe Kapitel 2.2): Neural Radiance Fields (NeRF) und ein Text-to-Image Diffusion Model namens „Imagen“ (Saharia et al., 2022). Imagen wird verwendet, um Bildvariationen eines eingegebenen Textes zu erstellen, die weiterverwendet werden, um ein Neural Radiance Field zu optimieren, wodurch eine dreidimensionale Repräsentation der Bilder entsteht (siehe Abb. 8)



Abbildung 8: Ergebnis Dream Diffusion (Poole et al., 2022, S.2)

In der 2023 erschienenen Publikation von Lin et al. „Magic3D“ (2023) verbessern die Autoren, die von Poole et al. vorgestellte „DreamFusion“ (2022) Methode, indem sie den Optimierungsprozess des NeRFs in zwei Schritte unterteilen. Die vorgestellte Methode schafft es so, Ergebnisse in halber Zeit zu produzieren und gleichzeitig die Qualität zu verbessern. „Our method, dubbed Magic3D, can create high quality 3D mesh models in 40 minutes, which is $2 \times$ faster than DreamFusion (reportedly taking 1.5 hours on average), while also achieving higher resolution. User studies show 61.7% raters to prefer our approach over DreamFusion“ (Lin et al., 2023 S.1).

In „Point-E“ stellen Nichole et al. (2022) eine noch schnellere Methode vor, die 3D Modelle innerhalb von ein bis zwei Minuten produzieren kann. Die vorgestellte Methode generiert zunächst ein einzelnes Bild mithilfe eines text-to-image Diffusion Models. Aus

diesem Bild generiert ein weiteres Diffusion Model eine Point Cloud und erschafft so eine dreidimensionale Repräsentation des im ersten Diffusion Model generierten Bildes.

Image-to-3D Modelle

Zur Image-to-3D Kategorie gehören Modelle wie „Make-it-3D“ (Tang et al., 2023), „GET3D“ (J. Gao et al., 2022) und „PIFuHD“ (Saito et al., 2020). Image-to-3D Modelle verwenden Bild- statt Textdaten als Input, um 3D Modelle zu generieren. Demnach gehört auch die im Kapitel 2.2 aufgeführte NeRF-Technologie, die eine Art „Foundation Technologie“ für viele weitere Frameworks bildet, zu diesen Technologien.

Mit „Make-it-3D“ präsentieren Tang et al. (2023) eine Methode zur hochpräzisen Erstellung von 3D Content aus einem einzigen Bild (siehe Abb. 9). „In this work, we investigate the problem of creating high-fidelity 3D content from only a single image“ (Tang et al., 2023). Die von den Autoren vorgestellte Methode ist, wie auch bei anderen bereits vorgestellten Papers, in zwei Stufen unterteilt (siehe Abb. 10). „In the first stage, we leverage diffusion prior to optimize a neural radiance field (NeRF)“ (Tang et al., 2023, S.2). Im ersten Schritt, der im Paper „Corse stage“ genannt wird, wird wie in der bereits vorgestellten Methode „DreamFusion“ (Lin et al., 2023) ein Diffusion Model verwendet, um ein NeRF (siehe Kapitel 2.2) zu optimieren. Anders als bei DreamFusion wird in Make-it-3D eine Bilddatei statt eines Textprompts verwendet, um das Diffusion Model anzuleiten. „We refine the model by leveraging the availability of ground-truth textures for regions that are observable in the reference image. To achieve this, we export the coarse NeRF model to textured point clouds and project reference textures onto their corresponding areas in the point clouds“ (Tang et al., 2023 S.2). Im zweiten Schritt, „Refine Stage“ genannt, wird das erstellte 3D Modell in eine Point Cloud umgewandelt und Stellen, die sich außerhalb des Blickwinkels des Referenzbildes befinden, werden textureiert.



Abbildung 9: Ergebnis Make-it-3D (Tang et al., 2023, S.16)

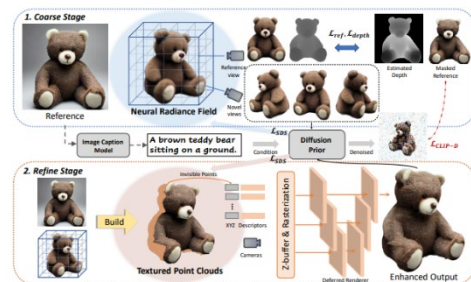


Abbildung 10: Überblick Architektur Pipeline (Tang et al., 2023, S.3)

Die Codes aller Publikationen sind frei öffentlich zugänglich und können von jedem verwendet werden. Allerdings ist die Verwendung dieser, wie auch bei den in Kapitel 3.4 vorgestellten Texture Modellen, mit teilweise sehr hohen Systemanforderungen sowie einem zumindest grundlegenden Verständnis von Programmiersprachen verknüpft. Deswegen haben sich viele Unternehmen dazu entschlossen, die angesprochenen Technologien leichter zugänglich zu machen. Ein Beispiel für eine aktuell sehr bekannte und beliebte Text-to-3D Anwendungen heißt „Genie“ und wird von Luma AI entwickelt

(Luma AI, o. D.). Genie ist zurzeit kostenlos über Discord nutzbar und ermöglicht es, aus einfachen Textprompts texturierte 3D Modelle zu generieren. Genie ist aber nicht die einzige KI-Applikation, die Luma AI zur Verfügung stellt. In ihrer gleichnamigen Smartphone-App wird Artists die Möglichkeit gegeben, die NeRF-Technologie zu nutzen. Objekte und ganze Räume können unter Anleitung über die Handy Camara aufgenommen und in NeRFs umgewandelt werden. Luma AI ist aber nicht der einzige Anbieter, der KI gestützte 3D Objekt Generation anbietet. Im Internet gibt es eine Vielzahl vergleichbarer Angebote für Text- sowie Image-to-3D (Masterpiece Studio, o. D.; 3DFY.ai, 2023; Common Sense Machines, Inc., o. D.).

3.6 Bearbeitung von 3D Modellen mit Hilfe von KI

In der durchgeführten Literaturrecherche wurden auch einige Frameworks identifiziert, die die Bearbeitung bereits bestehender 3D Modell durch KI erleichtern. Diese Technologien könnten im Arbeitsprozess des 3D Artists verwendet werden, um beispielsweise durch KI generierte 3D Modelle zu stilisieren und so zu einer einheitlichen Atmosphäre des Environments beizutragen.

Eines dieser Frameworks heißt „Instruct-NeRF2NeRF“. Mit „Instruct-NeRF2NeRF: Editing 3D Scenes with Instructions“ haben Haque et al. (2023) eine Möglichkeit vorgestellt NeRF-Szenen durch Textanweisungen zu bearbeiten (siehe Abb. 11). Bei der vorgestellten



Abbildung 9: Instruct NeRF2NeRF (Tang et al., 2023, S.1)

Methoden werden die der NeRF-Szene zugrundeliegenden Eingabebilder durch ein „image conditioned diffusion model (InstructPix2Pix)“ (Haque et al., 2023) iterativ bearbeitet, was zu einer NeRF-Szene führt, die Bearbeitungsanweisungen berücksichtigt. Im Klartext bedeutet das, dass die der NeRF-Szene zugrundeliegenden Eingabebilder mithilfe einer weiteren künstlichen Intelligenz (InstructPix2Pix) bearbeitet werden. InstructPix2Pix ermöglicht es Bilder mithilfe von Texteingabe zu bearbeiten (Brooks et al., 2023), gehört also demnach zu den bereits behandelten „Text-to-Image“ Modellen. Da das Neural Radiance Field durch diese Bilder trainiert wird, wird jede Veränderung der Trainingsbilder auch auf die volumetrische Szene des NeRFs übertragen, was dazu führt, dass man dieses über die Eingabe von Textbefehlen verändern kann.

Weitere Frameworks, die die KI basierte Bearbeitung von 3D Modellen ermöglichen, sind „SKED“ (Mikaeili et al., 2023-A) und „TextDeformer“ (W. Gao et al., 2023).

Mit SKED präsentieren Mikaeili et al. „A sketch-driven 3D editing technique designed to modify a pre-trained NeRF Scene“ (2023.-B, Min.1:20). Neben Zeichnungen des Nutzers nutzt SKED außerdem Textprompts als Input, um die NeRF-Szene zu bearbeiten. SKED erweitert also die KI gesteuerte Bearbeitung von NeRF-Scenes um die Funktionalität des Einzeichnens gewünschter Objekte (siehe Abb. 12).



Abbildung 10: SKED (Mikaeili et al., 2023, S.1)

TextDeformer bietet anders als die beiden bereits angeführten Frameworks, die Methoden zur Bearbeitung von NeRF-Szenen darstellen, eine Technik zur automatischen Bearbeitung von triangulierten Meshes durch Texteingabe (W. Gao et al., 2023, S1). Das Framework setzt dabei auf differenziertes Rendering (Kato et al., 2020), um Geometrien mit vortrainierten Bildencodern wie CLIP oder DINO zu verbinden.

3.7 KI Scene Generation

Weitere Einsatzmöglichkeiten von KI in der Gestaltung virtueller Umgebungen sind die sogenannten „Text Guided 3D Scene generation“ Models (Li et al., 2023, S.10). Diese schaffen es durch die Eingabe von Textprompts ganze 3D Scenes zu erstellen. Im Gegensatz zu anderen bereits vorgestellten KI-Tools, die einzelne Geometrien erstellen können, welche anschließend zu einer Umgebung zusammengesetzt werden müssen, erstellen diese Technologien ganze Umgebungen oder Räume auf Basis von Textprompts. Li et al. (2023, S10) gehen in ihrer Arbeit auf sechs verschiedene Ansätze ein, die die textbasierte Generation von 3D Scenes durch KI ermöglichen (Cohen-Bar et al., 2023; Po und Wetzstein, 2023; Fridman et al., 2023; Höllein et al., 2023; J. Zhang et al., 2023; Singer et al., 2023). Besonders stechen die Methoden „Text2Room“, (Höllein et al., 2023) und „Set-the-Scene“ (Cohen-Bar et al., 2023) hervor.

Wie der Name schon verrät, stellen die Autoren Höllein et al. (2023) mit „Text2Room“ eine Methode vor, die darauf spezialisiert ist, 3D Modelle von Räumen aus Textprompts zu generieren (siehe Abb. 13). Genauer gesagt stellen die Autoren eine Strategie vor, die Szenenbilder iterativ mit vorhandener Geometrie verschmilzt und so ein nahtlos texturiertes 3D Mesh eines Raumes erschafft. Dabei wird zunächst ein vorab trainiertes 2D-Text-zu-Bild Model verwendet, um eine Reihe an Bildern aus verschiedenen

Kamerapositionen zu synthetisieren. Um diese Bilder in eine 3D Szene umzuwandeln, kombinieren die Autoren monokulare Tiefenschätzung mit einem „text-conditioned inpainting model“ (Höllelein et al., 2023). Die monokulare Tiefenschätzung ist ein Prozess der Computer Vision, einem „Feld innerhalb der künstlichen Intelligenz (KI), das es Computern und System ermöglicht aussagefähige Informationen aus digitalen Bildern, Videos und anderen Visuellen Eingaben zu gewinnen“ (International Business Machines Corporation, o. D.), bei dem Tiefeninformationen eines Bilds vorhergesagt werden (Kim et al., 2023).

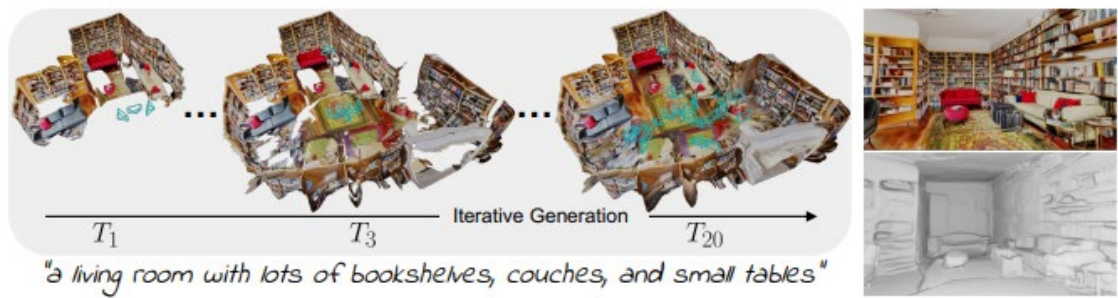


Abbildung 11: Text2Room (Höllelein et al., 2023, S.1)

Methoden wie die von Hollelein et al. (2023) haben nur eine begrenzte Kontrolle über die Platzierung oder das Erscheinungsbild der sich innerhalb der Szene befindlichen Objekte, die die Szene als ein Ganzes darstellen. In Szenarien, in denen man möglichst viel Kontrolle über die sich in der Szene befindenden Objekte haben möchte, wie z.B. im Kontext der 3D Environment Gestaltung, können diese Limitationen ein erhebliches Probleme darstellen.

Mit „Set-the-Scene“ stellen Cohen-Bar et al. (2023) eine Methode zur textbasierten Synthese von 3D Scenes vor, die diese Limitationen durch die Verwendung von „Object proxies“ umgeht. Ein Proxy repräsentiert die Platzierung eines Objekts innerhalb der Szene und definiert optional seine grobe Geometrie (Cohen-Bar et al., 2023, S.1). „The key to our approach is to represent each object as an independent NeRF“ (Cohen-Bar et al., 2023, S.1). In der vorgestellten Methode werden also die sich innerhalb der Szene befindenden Objekte als voneinander unabhängigen NeRFs dargestellt, wodurch eine



Abbildung 12: Set-the-Scene (Cohen-Bar et al. 2023, S.3)

Verschiebung der einzelnen Objekte, sowie eine Anpassung dieser möglich wird. Cohen-bar et al. (2023) optimieren abwechselnd jedes NeRF als einzelnes (local) und als Teil der gesamten Szene (global), wodurch eine vollständige Darstellung jedes Objekts erlernt wird und gleichzeitig eine harmonische Szene mit einheitlichem Stil und Lichtverhältnissen entsteht (siehe Abb. 14).

3.8 Large Language Models

Beim Arbeiten in 3D stoßen Artists immer wieder auf technische Probleme deren Lösungen nicht immer klar auf der Hand liegen. Zur Lösung dieser Probleme wenden sich Artists mittlerweile unter anderem an KIs, die auf Basis riesiger Datensätze trainiert werden und deswegen umfangreiches Wissen in vielen Themenbereichen haben. Damit sind textbasierte Modelle, wie beispielsweise GPT-4 gemeint, die „Large-Language Models“ oder auch „Transformer Netzwerke“ genannt werden. „LLMs beschränken sich bei Input und Output auf Text, können innerhalb dieses Formats aber sehr viele Stile abbilden, wie Konversationen, Programmcode in diversen Sprachen, Übersetzungen und diverse literarische Stile“ (Interviewpartner Zwei 2023, S.1).

Wofür können diese Modelle aber im Kontext der 3D Environment Gestaltung angewendet werden? Wie bereits am Anfang dieses Unterpunktes erwähnt, werden sie in der Lösung von technischen Problemen verwendet. „Ich sehe diese Werkzeuge grundsätzlich als Alternative zu Frage-Antwort-Websites wie Stack Overflow, die ich zum Lösen technischer Probleme nutze. Beide kommen in Verlässlichkeit und Vertrauenswürdigkeit nicht an wissenschaftliche Papers, Bücher und Online-Kurse heran, sind aber in der Regel die schnellste Lösung für kleine Probleme“ (Interviewpartner Zwei, 2023, S.2).

Eine weitere Anwendungsmöglichkeit von LLMs besteht in der Generierung von Code Elementen, denn die meisten 3D Programme bieten die Möglichkeit Code auszuführen. Besonders wenn komplexe Simulationen (z.B. spezielle Wasserbewegungen) erstellt werden sollen, wird Coding wichtig. Diesen Code kann man sich in LLMs wie ChatGPT schreiben lassen und innerhalb des Programms ausführen (Interviewpartner Zwei, 2023, Min.00:13:28).

4.0 Diskussion

4.1 Integration von Künstlicher Intelligenz in den Arbeitsprozess

Die Ergebnisse dieser Arbeit zeigen, dass es zum jetzigen Zeitpunkt (Januar 2024) eine Vielzahl an auf KI basierenden, Frameworks, Technologien und Workflows gibt, die von 3D Environment Artists genutzt werden können. Es wurden mehrere Teilbereiche der 3D Environment Gestaltung identifiziert, in denen KI aktuell eingesetzt werden kann. Dazu gehören direkte handwerkliche Aufgänge, wie das Texturieren, Generieren und Bearbeiten von 3D-Geometrien sowie kreative Aufgaben im Preprocessing Bereich, wie dem Erstellen von Referenzbildern, Moodboards oder auch Styleframes. Im Folgenden

werden die Ergebnisse der Arbeit hinsichtlich der Integrationsmöglichkeiten in den Arbeitsprozess von 3D Environment Artists diskutiert.

Ein 3D Environment ist nach der Definition von ProGlobalBusinessSolutions (2023) die dreidimensionale Repräsentation von geometrischen Daten, die der echten Welt nachempfunden sind. Diese Repräsentation der Daten findet über 3D Modelle statt, weswegen besonders die in Kapitel 3.5, 3.6 und 3.7 vorgestellten Tools zur Generation und Bearbeitung von 3D Modellen von besonderer Relevanz für 3D Environment Artists sind. Da die dabei generierten 3D Geometrien aktuell jedoch noch nicht den industriellen Standards entsprechen (Li et al. 2023, S.11), entscheiden sich Artists aktuell noch häufig gegen eine Verwendung dieser Tools (Interviewpartner Eins, 2023, Min.00:13:07). Die Modelle sind also aktuell noch nicht gut genug, um es tatsächlich auf die Leinwand oder in ein Videospiel zu schaffen. Sie könnten jedoch in frühen Phasen der Produktion, wie dem Blockout, als Platzhalter verwendet werden und so bereits früh ein besseres Bild des Endergebnisses kommunizieren. Hier wäre ein Plugin hilfreich, welches innerhalb von 3D Software, ausgewählte Geometrien auf Basis von Text bearbeitet.

Die meisten Tools die 3D Modelle generieren, sind jedoch noch sehr jung und befinden sich am Anfang ihrer Entwicklung. Betrachtet man die Entwicklungen anderer KI-Generatoren wie Midjourney (siehe Abb. 15), drängt sich die Vermutung auf, dass auch 3D Generatoren eine solche rapide Steigerung der Qualität durchmachen könnten. Um eine Entwicklung vergleichbar mit der von Midjourney beobachten zu können, müssten 3D Generatoren genau wie Bilddiffusionsmodelle auf riesigen Datensätzen aufgebaut werden. Für 3D Daten gibt es jedoch aktuell keine solchen Datensätze (Poole et al., 2022, S.1), wes-



Abbildung 13: The Evolution of Midjourney (Ma, 2023)

wegen 3D Generatoren alternative Ansätze, wie die Kombination von 2D Bildgeneratoren und NeRFs, zur Generation verfolgen. Deswegen kann auch nicht davon ausgegangen werden, dass diese eine solche beschleunigte Qualitätssteigerung haben werden, wie sie bei Bildgeneratoren wie Midjourney zu beobachten war. Dieser Antwortversuch vernachlässigt jedoch andere Faktoren, die einen Einfluss auf die Entwicklungen von 3D Generatoren haben können. Eine realistische Antwort auf diese Frage ist also nicht im Umfang dieser Arbeit möglich und könnte durch zukünftige Arbeiten genauer geklärt werden.

Auf Grund mangelnder Qualität entscheiden sich viele Artists gegen die Verwendung von 3D Generatoren und produzieren ihre 3D Modelle selbst. Wichtiger Aspekt in der Produktion von digitalen 3D Modellen ist die Texturierung, in der Bilddateien verwendet werden können. In Kapitel 3.4 werden verschiedene Diffusionsbasierte KI Tools vorgestellt, die die Generation von Bild Texturen ermöglichen und dabei auch größtenteils brauchbare Ergebnisse produzieren. Dabei übernehmen sie jedoch nur einen Teilaspekt der Texturierungsarbeit, da für die Verwendung von Bildtexturen zunächst eine UV Map erstellt werden muss. Dazu kommt, dass Modelle, die im Kontext der 3D Environment Gestaltung erstellt und texturiert werden, häufig aus mehreren Texturen bestehen, die miteinander kombiniert werden müssen, um die echte Welt nachbilden zu können. Hier könnte eine Applikation Abhilfe schaffen, die Texturen beispielsweise über einen Inpainting Modus auf dem 3D Modell miteinander verschmilzt und KI nutzt, um einen sinnvollen Übergang zwischen den Texturen zu schaffen.

Ein interessanter alternativer Ansatz zur Generation von dreidimensionalen Umgebungen, der ebenfalls einen Bildgenerator verwendet, wird in Kapitel 3.3 vorgestellt. Hier wird eine Depthmap einer ausgeblockten Szene zur Generierung eines Bildes verwendet. Diese Methode zeigt gut, wie KI aktuell neue Gestaltungsmöglichkeiten im Kontext der 3D Environment Gestaltung eröffnet, die dabei helfen, viel Zeit in der Produktion von Projekten einzusparen. Die vorgestellte Methode ist jedoch nur in Kontexten einsetzbar, in denen die Kameraperspektive vom Artist gewählt wird, da ein zu starke Kamerabewegung zur Verzerrung der Umgebung führt. Eine Anwendung zur Gestaltung von Umgebungen, in denen sich die Nutzer frei bewegen können, wie beispielsweise in Open World Videospielen, kommt hier also nicht in Frage. Geübte Augen erkennen außerdem schnell, dass in der Produktion der Umgebung KI verwendet worden ist, was jedoch durch weitere Verbesserung der Bildgeneratoren in Zukunft schwerer werden könnte und möglicherweise sogar als zeitgenössischer Stil angesehen werden kann.

Dreidimensionale Umgebungen dienen besonders in der Filmproduktion häufig als Hintergrund einer sich im Vordergrund abspielenden Szene. Aber auch innerhalb der 3D Environments werden Hintergründe in Form von 360 Grad Fotos verwendet (vgl. Kapitel 3.2). Besonders ist, dass diese 360 Grad Fotos auch dazu verwendet werden können, um die sich in der Szene befindende Geometrie zu beleuchten. Die Recherchen dieser Arbeit haben ergeben, dass es KI-Modelle gibt, die die Generation von 360 Grad Fotos ermöglichen (Blockade Labs, o. D.). Da im Zuge der Literaturrecherche keine weiteren KI basierten Methoden zu Beleuchtung von 3D Szenen gefunden wurden, stellen diese Modelle die erste KI-basierte Methode zu Beleuchtung von 3D Geometrie dar.

Zur Tätigkeit von 3D Environment Artists gehören auch Aufgaben wie die Erstellung von Moodboards und Referenzfotos. (vgl. Kapitel 3.1). Bildgeneratoren stellen hier aktuell

aber bereits eine sehr gute Alternative zu herkömmlichen Bildsuchmaschinen wie Pinterest dar, da die Ergebnisse über Texteingabe gut spezifizierbar sind (Interviewpartner Eins, 2023. Min.00:05.01). Aber auch in der Styleframing Phase, in der meist nur ein bestehender Stil auf das Projektthema übertragen wird, können diese Modelle dank der Imageprompt Funktion verwendet werden und so viel Zeit und überflüssige Arbeit einsparen (Interviewpartner Zwei, 2024, S.3). Es kann also gesagt werden, dass sich KI in diesem Bereich aktuell gut in den Arbeitsablauf integrieren lässt, um konzeptionelle Arbeit zu erleichtern.

Neben all diesen praxisbezogenen und meist mit KI-Bildgeneration verknüpften Anwendungsgebieten, gibt es eine auch Verwendung für KI basierte Sprachmodelle (Large Language Models) (vgl. Kapitel 3.8). Diese können zur Lösung von technischen Problemen verwendet werden oder das Schreiben von Code übernehmen (Interviewpartner Zwei, 2024, S.2). LLMs „kommen in Verlässlichkeit und Vertrauenswürdigkeit nicht an wissenschaftliche Papers, Bücher und Online-Kurse heran, sind aber in der Regel die schnellste Lösung für kleine Probleme“ (Interviewpartner Zwei, 2024, S.2). Aber auch hier kann davon ausgegangen werden das LLMs durch weiteres Training in Zukunft ein noch bessere Hilfe für spezielle technische Probleme leisten kann.

4.2 Ablösung des 3D Environment Artists durch Künstliche Intelligenz

Die Ergebnisse haben gezeigt, dass es zahlreiche auf KI basierende Methoden gib, die Aufgaben übernehmen können, die zu denen der 3D Environment Artists gehören. Die Qualität der Ergebnisse, die in einigen Bereichen dabei erzielt wird, entspricht zwar aktuell noch nicht der Qualität herkömmlicher Workflows, die Ergebnisse sind dafür aber bedeutend weniger aufwendig in der Produktion und in einem Bruchteil der Zeit präsentierbar. Es ist jedoch sicherlich nur eine Frage der Zeit, bis die Qualität hier für den Großteil der Produktionen ausreichend ist. Es wird deswegen unvermeidlich sein, das Potential von generativen KI-Modellen zu nutzen, wenn man als Artist wettbewerbsfähig bleiben möchte (Interviewpartner Eins, 2023, Min.00:30:12).

Die Aufgaben von 3D Artists in der Produktion sind jedoch nicht immer nur handwerklicher Natur. „Je nach Tätigkeit und Rolle im Team ist das Finden kreativer, innovativer Lösungen und Stile und das Führen des Kunden durch den kreativen Prozess mindestens genauso wichtig“ (Interviewpartner Zwei, 2023, S.4). Diffusionsbasierte KI-Modelle sind jedoch nicht in der Lage von sich aus Innovation zu schaffen. „Ihre Architektur ist darauf ausgelegt, die wahrscheinlichste - und damit generischste - Lösung für einen bestimmten Prompt zu finden. Wenn ich als Nutzer nicht in der Lage bin, einen

interessanten Prompt zu formulieren, bekomme ich auch keinen interessanten Output“ (Interviewpartner Zwei, 2023, S.4). Das lässt darauf schließen, dass in Zukunft Prompt Engineering ein wichtiger Faktor in der Arbeit von 3D Artists und damit natürlich auch 3D Environment Artists werden könnte. Diese These wird weiterhin durch Aussagen bestätigt wie „Dass KI Jobs abschafft, ist ein möglicher Ausgang, aber nicht der einzige: Man kann genauso davon ausgehen, dass KI einfach fester Teil unseres Workflows werden und sich die Aufgaben von 3D Artists mehr in Richtung der eines Art Directors verschieben“ (Interviewpartner Zwei, 2023, S.4).

Projekte im Grafikbereich sind in den meisten Fällen iterative Prozesse, die eine enge Zusammenarbeit zwischen Kunden und Studio bzw. Artist verlangt. Da Künstliche Intelligenz nicht in der Lage ist, selbst Expertise zu zeigen, den Kunden vor schlechten Entscheidungen zu warnen und Alternativen anzubieten, müssen Kunden, die diese Expertise wollen, weiterhin auf Menschen zugehen, die diese mitbringen (Interviewpartner Zwei, 2023, S.4).

Die Angst vor KI die aktuell in der Branche herrscht basiert möglicherweise auf einem verlorenen Fokus für die ursprüngliche Aufgabe: Visuelle Kommunikation. Es wurde sich zu lange auf komplexen Workflows ausgeruht, die zwar optisch ansprechende Bilder erzeugen, aber in vielen Fällen inhaltlich vollkommen uninteressant sind. KI droht aktuell diese Workflows zu vereinfachen, weswegen viele Artists Existenzängste bekommen. (Interviewpartner Zwei, 2023, S.5-6)

5.0 Fazit

Zum jetzigen Zeitpunkt (Januar 2024) gibt es eine Vielzahl an auf KI basierenden Workflows, Frameworks und Tools, die sich in den Arbeitsprozess von 3D Environment Artists integrieren lassen. In der Gestaltung von Umgebungen haben Artists eine große Menge an Aufgaben zu erfüllen, in welche sich KI aktuell unterschiedlich gut integrieren lässt. So gibt es derzeit zwar bereits KI-Tools zur Generierung von 3D Modellen, jedoch ist die Qualität der Modelle, die dabei erzeugt werden, noch nicht gut genug, um diese in großem Umfang in der Gestaltung von virtuellen Umgebungen nutzen zu können. In der Texturierung von bereits bestehenden Modellen, so wie in konzeptionellen Arbeiten, schaffen es KIs hingegen Ergebnisse zu erzeugen, die an die der herkömmlichen Arbeitsweise herankommen und diese teilweise sogar verbessern und erleichtern. Weiter ist auch der Kontext in dem ein Environment gestaltet wird ausschlaggebend, wenn es um die Beantwortung der Frage nach aktuellen Integrationsmöglichkeiten geht. Moderne Ansätze wie die von (Mickmumpitz, 2023) zeigen, dass Bildgeneratoren bereits einen großen Einfluss auf die Gestaltung von Umgebungen haben können, wenn diese in

Kontexten verwendet werden, in denen nur eine begrenzte Flexibilität der Kameraperspektive benötigt wird.

Doch die Aufgaben von 3D Artists sind nicht immer nur handwerklicher Natur. Auch das Beraten von Kunden und das Finden kreativer Lösungen ist Teil der gestalterischen Arbeit. KI ist jedoch nicht in der Lage, Kunden vor schlechten Entscheidungen zu bewahren oder von sich aus Innovation zu schaffen, weswegen sich Kunden, die diese Expertise brauchen, aktuell noch an Menschen wenden müssen, die diese mitbringen (Interviewpartner Zwei, 2023, S.4). Eine Ablösung des Artists ist somit aktuell noch unwahrscheinlich. Durch die immer breiter werdende Masse an KI Tools, die zu großen Teilen durch Eingabe von Text funktionieren, kann davon ausgegangen werden, dass in der Zukunft Prompt Engineering zu einem ausschlaggebenden Faktor in gestalterischer Arbeit wird und sich Jobs wie die der 3D Environment Artists immer mehr in die Richtung eines Artdirectors verschieben werden.

Literaturverzeichnis

- 3DFY.ai. (2023, 14. Juni). 3DFY.ai. Abgerufen am 4. Januar 2024, von <https://3dfy.ai/>
- Ahirwar, K. (2023, 19. Oktober). A very short introduction to diffusion models - Kailash Ahirwar - medium. *Medium*. <https://kailashahirwar.medium.com/a-very-short-introduction-to-diffusion-models-a84235e4e9ae>
- Bartle, R. A. (2003). *Designing virtual worlds*. New Riders. https://www.researchgate.net/publication/200025892_Designing_Virtual_Worlds
- Blender Market. (o. D.). Blender Market. Abgerufen am 22. Januar 2024, von https://blendermarket.com/search?utf8=%E2%9C%93&search%5Bq%5D=AI&search%5Bcategory_id%5D=&button=
- Blockade Labs. (o. D.). *Skybox AI*. Abgerufen am 12. Januar 2024, von <https://skybox.blockadelabs.com/>
- Brems, M. (o. D.). *A beginner's guide to the CLIP model - KDNuggets*. KDNuggets. Abgerufen am 20. Dezember 2023, von <https://www.kdnuggets.com/2021/03/beginners-guide-clip-model.html>
- Brooks, T., Holynski, A. & Efros, A. A. (2023). InstructPix2Pix: Learning to Follow Image Editing Instructions. *arXiv*. <https://doi.org/10.1109/cvpr52729.2023.01764>
- Carson-Katri. (2023, 7. November). *GitHub - Carson-katri/dream-textures: Stable diffusion built-in to blender*. GitHub. Abgerufen am 2. Januar 2024, von <https://github.com/carson-katri/dream-textures>
- Chen, D. Z., Siddiqui, Y., Lee, H., Tulyakov, S. & Nießner, M. (2023). Text2Tex: Text-driven texture synthesis via diffusion models. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2303.11396>
- Chen, R., Chen, Y., Jiao, N. & Jia, K. (2023). Fantasia3D: Disentangling geometry and appearance for high-quality Text-to-3D content creation. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2303.13873>
- Chen, Y., Chen, R., Lei, J., Zhang, Y. & Jia, K. (2022). TANGO: text-driven photorealistic and robust 3D stylization via lighting decomposition. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2210.11277>
- Cohen-Bar, D., Richardson, E., Metzger, G., Giryas, R. & Cohen-Or, D. (2023). Set-the-Scene: Global-Local Training for Generating Controllable NeRF Scenes. *arXiv*. <https://doi.org/10.48550/arXiv.2303.13450>
- Common Sense Machines, Inc. (o. D.). *CSM — Turn photos and videos into 3D worlds*. Abgerufen am 4. Januar 2024, von <https://www.csm.ai/>
- Danthree Studio. (2024, 3. Januar). Was ist ein 3D Mesh Modell? (Definition & Beispiele). *DANTHREE*. <https://www.danthree.studio/blog-cgi/was-ist-ein-3d-mesh-modell-definition-beispiele>
- Dawood, A. (2023, 7. Juli). *Number of midjourney users and statistics*. MLYearning. <https://www.mlyearning.org/midjourney-users-statistics/>
- Digital Synopsis. (2023, 31. Januar). *46 Famous movie scenes before and after visual effects*. <https://digitalsynopsis.com/design/movies-before-after-green-screen-cgi/>
- Fridman, R., Abecasis, A., Kasten, Y. & Dekel, T. (2023). SceneScape: Text-Driven Consistent Scene Generation. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2302.01133>
- Gao, J., Shen, T., Wang, Z., Chen, W., Yin, K., Li, D., Litany, O., Gojčić, Ž. & Fidler, S. (2022). GET3D: a generative model of high quality 3D textured shapes learned from images. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2209.11163>

- Gao, W., Aigerman, N., Groueix, T., Kim, V. G. & Hanocka, R. (2023). TextDeformer: Geometry manipulation using text guidance. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2304.13348>
- Goulekas, K. E. (2001). Visual effects in a digital world. In *Elsevier eBooks*. <https://doi.org/10.1016/b978-0-12-293785-9.x5127-9>
- Haque, A., Tancik, M., Efros, A. A., Holynski, A. & Kanazawa, A. (2023). Instruct-NeRF2NeRF: Editing 3D scenes with instructions. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2303.12789>
- Hatzius, J., Briggs, J., Kodnani, D. & Pierdomenico, G. (2023, 20. April). *Künstliche Intelligenz – die Folgen für die Wirtschaft*. gs.de. Abgerufen am 22. November 2023, von <https://www.gs.de/de/articles/kuenstliche-intelligenz-die-folgen-fuer-die-wirtschaft>
- Higgins, S. (2021, 23. Februar). Alles, was Sie über Punktwolken wissen müssen | NAVVIS. *NavVis GmbH*. <https://www.navvis.com/de/blog/everything-you-need-to-know-about-point-clouds-navvis>
- Höllein, L., Cao, A., Owens, A., Johnson, J. C. & Nießner, M. (2023). Text2Room: Extracting textured 3D meshes from 2D Text-to-Image models. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2303.11989>
- International Business Machines Corporation. (o. D.). *What is Computer Vision?* IBM. Abgerufen am 15. Januar 2024, von <https://www.ibm.com/de-de/topics/computer-vision>
- Kato, H., Beker, D., Morariu, M. A., Ando, T., Matsuoka, T., Kehl, W. & Gaidon, A. (2020). Differentiable rendering: a survey. *arXiv (Cornell University)*. <https://arxiv.org/pdf/2006.12057.pdf>
- Kim, I. S., Kim, H., Lee, S. & Jung, S. K. (2023). HeightNet: Monocular object height Estimation. *Electronics*, 12(2), 350. <https://doi.org/10.3390/electronics12020350>
- Leonardo Interactive Pty Ltd. (o. D.). *AI Art Generator - Create art, images & more | Leonardo AI*. AI Art Generator - Create Art, Images & More | Leonardo AI. Abgerufen am 2. Januar 2024, von <https://leonardo.ai/>
- Li, C., Zhang, C., Waghvase, A., Lee, L., Rameau, F., Yang, Y., Bae, S. & Hong, C. S. (2023). Generative AI Meets 3D: A survey on Text-to-3D in AIGC era. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2305.06131>
- Lin, C. (2023, 25. März). *Magic3D: High-Resolution Text-to-3D content Creation*. *arXiv.org*. <https://arxiv.org/abs/2211.10440>
- Luma AI. (o. D.). *Genie, a research preview of Luma's generative 3D foundation model*. Genie, a research preview of Luma's generative 3d foundation model. Abgerufen am 4. Januar 2024, von <https://lumalabs.ai/genie>
- Ma, Y. (2023, 28. Dezember). *All Midjourney versions (V1-V6) compared: The evolution of Midjourney*. Aituts. <https://aituts.com/midjourney-versions/>
- Masterpiece Studio. (o. D.). Masterpiece Studio. Abgerufen am 5. Januar 2024, von <https://masterpiecestudio.com/>
- Mickmumpitz. (2023, 15. Oktober). *Use AI to create amazing 3D animations! [Blender + stable diffusion]* [Video]. YouTube. <https://www.youtube.com/watch?v=YW3FzfVxAEs>
- Midjourney tile parameter. (o. D.). Abgerufen am 2. Januar 2024, von <https://docs.midjourney.com/docs/tile>
- Mikaeili, A., Perel, O., Cohen-Or, D. & Mahdavi-Amiri, A. (2023-A). SKED: Sketch-guided text-based 3D editing. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2303.10735>
- Mikaeili, A., Perel, O., Safaee, M., Cohen-or, D. & Mahdavi-Amiri, A. (2023-B). *SKED: Sketch-guided Text-based 3D Editing* [Video]. sked-

- paper.github.io. Abgerufen am 15. Januar 2024, von <https://sked-paper.github.io/>
- Muehmel, K. (2023, 7. Juni). *What is a large language model, the tech behind ChatGPT?* dataiku. <https://blog.dataiku.com/large-language-model-chatgpt>
- Nichol, A., Jun, H., Dhariwal, P., Mishkin, P. & Chen, M. (2022). Point-E: a system for generating 3D point clouds from complex prompts. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2212.08751>
- NVIDIA Corporation. (o. D.-A). *NVIDIA Canvas App: AI-Powered Painting*. NVIDIA. Abgerufen am 12. Januar 2024, von <https://www.nvidia.com/en-us/studio/canvas/>
- NVIDIA Corporation. (o. D.-B). *Publications*. NVIDIA Toronto AI Lab. Abgerufen am 3. November 2023, von <https://research.nvidia.com/labs/toronto-ai/>
- Oppenlaender, J. (2023). A taxonomy of prompt modifiers for text-to-image generation. *Behaviour & Information Technology*, 1–14. <https://doi.org/10.1080/0144929x.2023.2286532>
- OPUSAI. (o. D.). Abgerufen am 30. Oktober 2023, von <https://opus.ai/>
- Photogrammetrie-Software – Fotos in 3D-Scans | AutoDesk*. (o. D.). Abgerufen am 3. Januar 2024, von <https://www.autodesk.de/solutions/photogrammetry-software#:~:text=Was%20ist%20Photogrammetrie%3F,Zeichnungen%20oder%203D%2DModelle%20umgewandelt.>
- Po, R. & Wetzstein, G. (2023). Compositional 3D scene generation using locally conditioned diffusion. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2303.12218>
- Poole, B., Jain, A., Barron, J. T. & Mildenhall, B. (2022). DreamFusion: Text-to-3D using 2D diffusion. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2209.14988>
- ProGlobalBusinessSolutions. (2023, 27. Oktober). *What is 3D Environment Modeling? Guide for 3D Artists - PGBS*. PGBS. <https://www.proglobal-businesssolutions.com/what-is-3d-environment-modeling/#:~:text=A%203D%20environment%20is%20a,that%20give%20you%20visual%20sensations.>
- Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G. & Sutskever, I. (2021). Learning Transferable Visual Models From Natural Language Supervision. *arXiv (Cornell University)*. <https://arxiv.org/abs/2103.00020>
- Raveling, J. (2022, 3. Mai). *Was ist ein neuronales Netz?* WFB-Bremen. <https://www.wfb-bremen.de/de/page/stories/digitalisierung-industrie40/was-ist-ein-neuronales-netz>
- Richardson, E., Metzger, G., Alaluf, Y., Giryas, R. & Cohen–Or, D. (2023). Texture: Text-Guided Texturing of 3D Shapes. *arXiv*. <https://doi.org/10.1145/3588432.3591503>
- Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E., Ghasemipour, S. K. S., Ayan, B. K., Mahdavi, S., Lopes, R. G., Salimans, T., Ho, J., Fleet, D. J. & Norouzi, M. (2022). Photorealistic Text-to-Image diffusion models with deep language understanding. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2205.11487>
- Saito, S., Simon, T., Saragih, J. & Joo, H. (2020). PIFuHD: Multi-Level Pixel-Aligned Implicit Function for High-Resolution 3D Human Digitization. *arXiv*. <https://doi.org/10.1109/cvpr42600.2020.00016>
- Schroeder, R. (2008). Defining Virtual Worlds and Virtual Environments. *Journal of Virtual Worlds Research*, 1(1). <https://doi.org/10.4101/jvwr.v1i1.294>

- Singer, U., Sheynin, S., Polyak, A., Ashual, O., Makarov, I., Kokkinos, F., Goyal, N., Vedaldi, A., Parikh, D., Johnson, J. C. & Taigman, Y. (2023). Text-To-4D Dynamic Scene Generation. *arXiv (Cornell University)*.
<https://doi.org/10.48550/arxiv.2301.11280>
- Tang, J., Wang, T., Zhang, B., Zhang, T., Yi, R., Ma, L. & Chen, D. (2023). Make-It-3D: High-Fidelity 3D creation from a single image with diffusion prior. *arXiv (Cornell University)*.
<https://doi.org/10.48550/arxiv.2303.14184>
- Unreal Engine Marketplace. (o. D.). Unreal Engine. Abgerufen am 30. Oktober 2023, von <https://www.unrealengine.com/marketplace/en-US/assets?count=20&sortBy=effectiveDate&sortDir=DESC&start=0&tag=2>
- Was ist 3D-Texturierung? | Adobe Substance 3D. (o. D.). Abgerufen am 2. Januar 2024, von <https://www.adobe.com/de/products/substance3d/discover/3d-texturing.html#:~:text=Mit%20Texturen%20lassen%20sich%20einfache,Gras%2C%20Kies%20oder%20Stein%20simulieren.>
- Was ist Künstliche Intelligenz und wie wird sie genutzt? (2020, 14. September). European Parlament. Abgerufen am 11. November 2023, von <https://www.europarl.europa.eu/news/de/headlines/society/20200827STO85804/was-ist-kunstliche-intelligenz-und-wie-wird-sie-genutzt#:~:text=K%C3%BCnstliche%20Intelligenz%20ist%20die%20F%C3%A4higkeit,ein%20bestimmtes%20Ziel%20zu%20erreichen.>
- Wennker, P. (2020). *Künstliche Intelligenz in der Praxis: Anwendung in Unternehmen und Branchen: KI wettbewerbs- und zukunftsorientiert einsetzen*. Springer Gabler.
- Zhang, J., Li, X., Wan, Z., Wang, C. & Liao, J. (2023). Text2NERF: Text-Driven 3D scene generation with neural radiance fields. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2305.11588>
- Zhang, L. & Agrawala, M. (2023). Adding conditional control to Text-to-Image diffusion models. *arXiv (Cornell University)*.
<https://doi.org/10.48550/arxiv.2302.05543>

Abbildungsverzeichnis

Abbildung 1: Neural Radiance Field (Mildenhall et al. 2020, S.2)..... 7

Abbildung 2: 3D Environment als Hintergrund (Digital Synopsis, 2023) 9

Abbildung 3: Sky-Dome (in Anlehnung an Blockade Labs, o. D) 9

Abbildung 4: 360 Grad Foto generiert durch Skybox AI (Blockade Labs, o.D)..... 10

Abbildung 5: Blockout, Depthmap & AI Bild (Mickmumpitz, 2023)..... 10

Abbildung 6: Verzerrte Texturen (Mickmumpitz, 2023, Min. 08:08)..... 11

Abbildung 7: Seamless und nicht seamless Texturen generiert mit Dall E (oben) und Midjourney (unten) (eigene Darstellung mit Elementen generiert durch Midjourney und Dall-E) 12

Abbildung 8: Ergebnis Dream Diffusion (Poole et al., 2022, S.2)..... 13

Abbildung 9: Ergebnis Make-it-3D (Tang et al., 2023, S.16)..... 14

Abbildung 10: Überblick Architektur Pipeline (Tang et al., 2023, S.3)..... 14

Abbildung 11: Instruct NeRF2NeRF (Tang et al., 2023, S.1) 15

Abbildung 12: SKED (Mikaeili et al., 2023, S.1) 16

Abbildung 13: Text2Room (Höllelein et al., 2023, S.1) 17

Abbildung 14: Set-the-Scene (Cohen-Bar et al. 2023, S.3) 17

Abbildung 15: The Evolution of Midjourney (Ma, 2023)..... 19

KI Verzeichnis

Verwendete KI-Anwendungen:

- Dall-E
- Skybox AI (Blockade Labs. o.D.)
- Midjourney

Einzelner Prompt zur Generierung des in Abbildung 3 & 4 verwendeten 360 grad Fotos – **Dokumentation finaler Prompt:**

Das in Abb. 3 und Abb.4 zu sehende 360 Grad Foto wurde durch das Tool Skybox AI generiert.

- Prompt Skybox AI
 - „Mountains, morning light“(Style – Realism)

Einzelne Prompts zur Generierung der in Abbildung 7 zu sehenden Bildtexturen – **Dokumentation finaler Prompts:**

Die in Abb. 7 zu sehenden bildtexturen wurden durch die Tools Midjourney und Dall-E generiert.

- Prompt Dall E:
 - „Create a Seamless texture of a red Brickwall“
- Prompt Midjourney:
 - "Create a Seamless texture of a red Brickwall --tile"



Eidesstattliche Erklärung

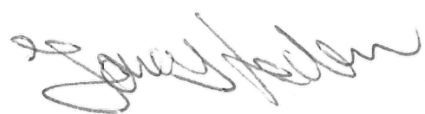
Ich, Jonas Lederer

geboren am 08. Mai 2000

erkläre hiermit, die vorliegende Arbeit selbständig und ohne fremde Hilfe angefertigt zu haben. Dabei habe ich mich keiner anderen Hilfsmittel bedient als derjenigen, die im beigefügten Quellenverzeichnis genannt sind.

Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind von mir als solche kenntlich gemacht.

München, den 26. Januar 2024
Studienort


.....
Unterschrift Studierende/r (= Verfasser/in)