



Planung und Entwicklung eines KI-Modells zum Transformieren von handschriftlichen Zeichnungen in fotorealistische Bilder

Robin Rausch und Florian Maslowski

2. November 2023

Matrikelnummer

Kurs

Ausbildungsfirmen

Ausbildungsorte

Betreuer

RR: 4718220 und FM: 2737788

TINF21B

Dürr Dental SE und Dürr Systems AG

Bietigheim-Bissingen

Prof. Dr.-Ing. Dipl. Inf. Kötter

Abstract

Deutsch

Die Forschung im Bereich der künstlichen Intelligenz umfasst viele verschiedene Bereiche der Informationstechnik und Mathematik. Aufgrund der aktuellen Entwicklungen und stetig fortschreitenden Forschung, ergeben sich immer mehr Möglichkeiten, diese Technologie anderswo einzusetzen. Die Arbeit beschäftigt sich mit der Planung und Entwicklung eines Modells basierend auf künstlicher Intelligenz, welches in der Kategorie Bild-zu-Bild einzuordnen ist. Ziel der Arbeit ist ein funktionsfähiges Modell, das handgefertigte Zeichnungen und Skizzen in fotorealistische Bilder umwandeln kann. Die Architektur dieses Modells soll aus bereits bestehenden Technologien konstruiert werden (Beispielsweise "Generative Adversarial Networks"). Das Modell soll zur Kreativität und Inspiration bei Designprozessen verwendet werden können. Ergebnis dieser Arbeit soll eine Applikation oder Website mit benutzerfreundlicher Oberfläche sein, welche das Erzeugen fotorealistischer Bilder anhand Skizzen ermöglicht. Benutzer*innen können dann über eine auf der Oberfläche angezeigte Zeichenfläche Skizzen erstellen. Diese werden dann in das zu erstellende Modell gegeben und das Ergebnis des Modells wird anschließend dem/der Benutzer*in angezeigt und zum Herunterladen angeboten. Trainiert wird das Modell über ein zu erstellendes Datenset, welches aus bestehenden Daten aus dem Internet und selbst erzeugten Daten zusammengesetzt wurde.

Begonnen wird mit der Erstellung eines Datensatzes. Hierzu werden Bilder aufgenommen und folglich als handgemalte Skizzen erstellt. Zusätzlich werden Open-Source-Datensets aus dem Web verglichen und teilweise oder ganz übernommen. Die Bilder werden dann in einem Datensatz zusammengefasst und später für das Trainieren des Modells verwendet. Nebenher erfolgt eine ausführliche Recherche zur Thematik und speziell der verschiedenen Möglichkeiten der Architektur des Modells. Darauf aufbauend soll die Struktur und allgemeine Planung des Modells erfolgen. Nach Abschluss der Planung und Vorbereitung wird mit der Entwicklung des Modells begonnen. Diese wird voraussichtlich in der Programmiersprache Python umgesetzt werden. Zuletzt erfolgt die Implementierung des Modells in eine Applikation. Zusätzlich wird eine

Oberfläche erstellt, auf welcher Benutzer*innen Skizzen erstellen können und an das Modell gegeben werden. Das resultierende Bild soll dann wieder ausgegeben werden.

Englisch

The research in the field of artificial intelligence encompasses many different areas of information technology and mathematics. Due to current developments and ongoing research, more and more opportunities are emerging to apply this technology elsewhere. The work deals with the planning and development of a model based on artificial intelligence, which falls into the category of image-to-image. The goal of the work is to create a functional model that can convert hand-drawn drawings and sketches into photorealistic images. The architecture of this model is intended to be constructed from existing technologies, such as "Generative Adversarial Networks." The model should be usable for fostering creativity and inspiration in design processes. The outcome of this work is expected to be an application or website with a user-friendly interface that allows the generation of photorealistic images based on sketches. Users can create sketches on a drawing area displayed on the interface. These sketches are then input into the model to generate results, which are subsequently displayed to the user and made available for download. The model will be trained using a dataset to be created, composed of existing data from the internet and self-generated data. The process begins with the creation of a dataset, where images are captured and subsequently transformed into hand-drawn sketches. Additionally, open-source datasets from the web are compared and may be partially or fully incorporated. The images are then compiled into a dataset and used for training the model. Simultaneously, comprehensive research on the subject matter and, specifically, the various architectural possibilities of the model is conducted. Building upon this, the structure and general planning of the model will be developed. After completing the planning and preparation, the model's development phase begins, which is likely to be implemented in the Python programming language. Finally, the model is implemented into an application, and a user interface is created to allow users to create sketches and input them into the model. The resulting image is then displayed again.

Vorwort

Diese Arbeit entstand im Prozess des Absolvierens meines Dualen Studiums und handelt über ...

Ehrenwörtliche Erklärung

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Aus den benutzten Quellen, direkt oder indirekt, übernommene Gedanken habe ich als solche kenntlich gemacht. Diese Arbeit wurde bisher in gleicher oder ähnlicher Form oder auszugsweise noch keiner anderen Prüfungsbehörde vorgelegt und auch nicht veröffentlicht.

2. November 2023, Bietigheim-Bissingen

Datum, Ort

Unterschrift

Unterschrift

Inhaltsverzeichnis

Abkürzungsverzeichnis	VI
1 Einführung	1
2 Motivation	2
3 Grundlagen	3
3.1 Künstliche Intelligenz	3
3.2 Maschinelles Lernen	3
3.3 Deep Learning	4
3.4 Generative Modelle	6
3.4.1 Generative Adversarial Networks (GAN)	6
3.4.2 Diffusion	6
3.5 Historie	6
4 Problemstellung	6
5 Evaluation generativer Modelle	7
6 Konzipierung	7
7 Implementierung	7
7.1 Datensatz erstellen	7
7.2 Entwickeln eines generativen Modells	7
7.3 Aufsetzen der Infrastruktur	7
8 Herausforderungen und Hindernisse	7
9 Kritische Reflexion und Ausblick	7
Literatur	8

Abkürzungsverzeichnis

vgl. vergleiche

KI Künstliche Intelligenz

GAN Generative Adversarial Networks

KI Künstliche Intelligenz

GPU Graphics Processing Unit

CNN Convolutional Neural Networks

Abbildungsverzeichnis

1	Deep Learning Modell [2]	5
2	Evaluationstabelle generativer Modelle	7

1 Einführung

In der heutigen Ära der künstlichen Intelligenz hat sich die Computerbildgebung zu einem dynamischen und faszinierenden Forschungsgebiet entwickelt. Die Entwicklung von Algorithmen und Modellen zur Generierung von realistischen Bildinhalten aus unterschiedlichen Quellen ist zu einer bemerkenswerten Herausforderung geworden. Dies ist insbesondere im Kontext von Handzeichnungen und Skizzen von großer Bedeutung. "MAGIC" steht für "Marvelous AI for Generating Image Content" und repräsentiert einen wichtigen Fortschritt in diesem Bereich.

Die Fähigkeit, aus handgezeichneten Skizzen fotorealistische Bilder zu erzeugen, ist nicht nur von technischem Interesse, sondern hat auch eine breite Palette von Anwendungen in Kunst, Design, Animation und sogar im Bereich der Medizin. Diese Studienarbeit widmet sich der Untersuchung und Analyse von MAGIC und seiner Rolle bei der Umwandlung von Skizzen in eindrucksvolle und realistische Bildinhalte.

Im Verlauf dieser Arbeit werden wir einen detaillierten Blick auf die Funktionsweise von MAGIC werfen, die zugrundeliegenden Technologien und Algorithmen erforschen und verschiedene Anwendungsfälle beleuchten. Darüber hinaus werden wir die Herausforderungen und Möglichkeiten in diesem Bereich erörtern, um ein umfassendes Verständnis für diese spannende Forschungsrichtung zu vermitteln.

2 Motivation

...

3 Grundlagen

3.1 Künstliche Intelligenz

Es ist schwer Künstliche Intelligenz (KI) genau zu definieren. Eine der verbreitetsten Definitionen geht aus der Arbeit "Computing machinery and intelligence" von Alan Turing hervor. Alan Turing war ein britischer Logiker, Mathematiker sowie Informatiker und Kryptoanalytiker. In dieser Arbeit führte er den bekannten Turing Test durch, in welchem eine Maschine als intelligent eingestuft wird, wenn sie in einer Konversation nicht von einem Menschen unterschieden werden kann. [vergleiche (vgl.) 1].

3.2 Maschinelles Lernen

Maschinelles Lernen bezieht sich auf Algorithmen und statistische Modelle, welche anhand eines Trainingsdatensatzes bestimmte Muster erkennen oder Schlüsse ziehen. Um ein solches Modell zu trainieren, wird ein Teil des Trainingsdatensatzes nicht zum Training sondern zum Testen verwendet wird. [vgl. 1, S.424].

Bei maschinellem Lernen wird zwischen verschiedenen Verfahren differenziert:

Überwachtes Lernen: Beim überwachten Lernen werden Algorithmen mit ausgewählten Trainingsdaten versorgt, die aus Eingabedaten und den zugehörigen Ausgabedaten bestehen. Das Modell lernt, eine Assoziierung von Eingabe zu Ausgabe herzustellen. Dies wird häufig für Aufgaben wie Klassifikation und Regression verwendet.

Unüberwachtes Lernen: Im unüberwachten Lernen werden Algorithmen mit Daten ohne vorherige Zuordnung oder Kennzeichnung trainiert. Ziel ist es, verborgene Strukturen oder Muster in den Daten zu erkennen. Dazu werden beispielsweise Clusterbildungen oder Dimensionsreduktionen verwendet.

Verstärktes Lernen: Beim verstärkenden Lernen lernt ein Modell durch Interaktion mit seiner Umgebung. Es erhält Feedback in Form von Belohnungen oder Bestrafungen und passt seine Aktionen

an, um die Gesamtbelohnung zu maximieren. Dieses Paradigma wird in Bereichen wie Robotik und Spieltheorie eingesetzt.

Für das entwickeln eines intelligenten Modells ist Training sowie Testen essenziell. W3schools empfiehlt dafür den Datensatz in Trainings- und Testdaten aufzuteilen. Das Verhältnis wird mit 80% für Trainingsdaten und 20% für Testdaten angegeben [vgl. 4].

Anwendungen:

Das maschinelle Lernen findet in breitgefächerten Gebieten Anwendung. Aktuell wird es dazu verwendet, um in der Medizin neue Arzneimittel, Impfstoffe, etc. zu erschaffen. Außerdem wird es in Laboren für die Diagnose von Krankheiten und Viren verwendet. In der Wirtschaft wird maschinelles Lernen für die Risikobewertung, Aktienhandel, etc. verwendet. Ein Anwendungsgebiet, welches für diese Arbeit hohe Relevanz ist, ist die Bildverarbeitung. Dort wird maschinelles Lernen für Objekterkennung, Bildtransformationen, etc. verwendet.

3.3 Deep Learning

Als Teilgebiet des maschinellen Lernens, wird bei dieser Art des Lernens auf den Nachbau eines menschlichen Gehirns gesetzt. Im menschlichen Gehirn befinden sich Neuronen/Gehirnzellen, welche miteinander verbunden sind. Beim Anlernen neuen Wissens bilden sich neue Neuronenbindungen. Diese Verbindungen werden durch öftere Verwendungen gestärkt. Auf der anderen Seite werden selten genutzte Verbindungen schwächer, bis sie nicht mehr existent sind. Deep Learning Modelle versuchen dieses Prinzip nachzubauen. Dabei werden Neuronen durch Knotenpunkte repräsentiert, welche in mehreren Schichten/Layer angeordnet sind. Während dem Training werden die Gewichtungen der verschiedenen Verbindungen/Kanten mit dem Backpropagation-Prozess optimiert. Neuronale Netzwerke, welche über mehrere "Hidden Layers" verfügen, werden als "deep" neuronale Netze bezeichnet. [vgl. 1, S.424]

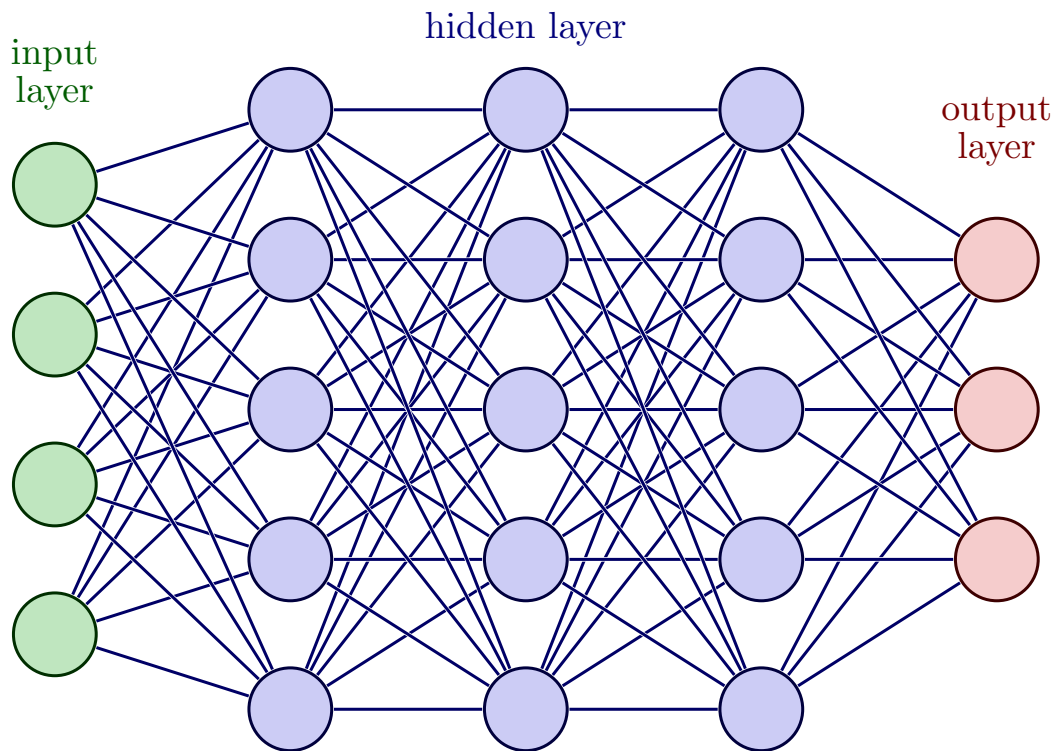


Abbildung 1: Deep Learning Modell [2]

Das Konzept des Deep Learnings war schon im Jahr 1943 bekannt. Ausschlaggebend dafür war der von Warren McCullough und Walter Pitts geschriebene Artikel "A Logical Calculus of Ideas Immanent in Nervous Activity". In diesem Artikel wurde erstmals das mathematische Modell eines neuronalen Netzwerks ausgearbeitet. 1950 wurde "Snarc", der erste Computer mit neuronalem Netzwerk, von den Harvard-Studenten Marvin Minsky und Dean Edmonds erschaffen. Im selben Jahr hat auch Turing seine Arbeit "Computing machinery and intelligence", welche den Turing-Test beinhaltet, veröffentlicht [vgl. 3]. Jedoch war damals die Rechenleistung nicht ausreichend effizient und bezahlbar, um im Alltag Fuß zu fassen. Im Jahr 2013 wurde erkannt, dass Grafikkarten/Graphics Processing Unit (GPU), welche eigentlich für die Berechnung 3-Dimensionaler Grafiken in Computerspielen konzipiert waren, große Anwendung im repetitiven Training von neuronalen Netzwerken finden können.

Convolutional Neural Networks (CNN)

CNN sind spezielle neuronale Netze, dessen Architektur besonders in der Bildverarbeitung als effektiv gilt. In den letzten Jahren stieg die Anzahl an Anwendungen aufgrund der demonstrierbaren Effizienz und Verfügbarkeit. Unternehmen wie Google und Facebook investierten in die Entwicklung und Forschung dieser Netze. Daraus resultierten Architekturen, wie zum Beispiel Inception von Google und ResNet von Microsoft. Diese Architekturen können über Frameworks, wie TensorFlow und PyTorch für eigene Zwecke verwendet und weiterentwickelt werden. Ein häufiger Ansatz ist es vortrainierte Architekturen zu modifizieren, um individuelle Anwendungen zu meistern.

2017 wurde ein Meilenstein in der Medizin erreicht. Die Zuverlässigkeit eines CNN konnte erstmals mit der eines Dermatologen verglichen werden. Für dieses Netzwerk wurde ein vortrainiertes Modell von Google für die medizinische Anwendung optimiert und mit einem eigenen Datensatz trainiert [vgl. 1].

3.4 Generative Modelle

Generative Modelle zeichnen sich dadurch aus, Daten zu produzieren/generieren. Dabei wird anhand des Inputs, ein "passender" Output generiert. Die Qualität des Outputs hängt von mehreren Faktoren ab. Einige davon sind:

Datenset:

Verwendete Algorithmen

...

3.4.1 Generative Adversarial Networks (GAN)

3.4.2 Diffusion

3.5 Historie

4 Problemstellung

...

5 Evaluation generativer Modelle

Modell	Bildqualität	Vielseitigkeit	Trainingszeit	Anwendungen
GAN	Hoch	Vielseitig	Lang	Bildgenerierung
Pix2Pix	Hoch	Abhängig	Mittel	Bild-zu-Bild-Übersetzung
CycleGAN	Gut	Hoch	Mittel	Bild-zu-Bild-Übersetzung
VAE	Mittel	Mittel	Kurz	Bildgenerierung
SRGAN	Sehr hoch	Abhängig	Lang	Super-Resolution
Neural Style Transfer	Künstlerisch	Niedrig	Kurz	Stilübertragung
Diffusion Model	Hoch	Hoch	Lang	Vielseitige Anwendungen

Abbildung 2: Evaluationstabelle generativer Modelle

6 Konzipierung

7 Implementierung

7.1 Datensatz erstellen

7.2 Entwickeln eines generativen Modells

7.3 Aufsetzen der Infrastruktur

8 Herausforderungen und Hindernisse

9 Kritische Reflexion und Ausblick

Literatur

- [1] X. Du-Harpur u. a. "What is AI? Applications of artificial intelligence to dermatology". In: *British Journal of Dermatology* 183.3 (Sep. 2020), S. 423–430. ISSN: 0007-0963. DOI: 10.1111/bjd.18880. eprint: <https://academic.oup.com/bjd/article-pdf/183/3/423/46341283/bjd0423.pdf>. URL: <https://doi.org/10.1111/bjd.18880>.
- [2] Izaak Neutelings. *Neural Networks*. Einsichtnahme: (11.01.2023). 2023. URL: https://tikz.net/wp-content/uploads/2022/01/neural_networks.tex.
- [3] Data Science. *Künstliche Intelligenz: Definition, Geschichte, Chancen und Risiken*. Einsichtnahme: (12.01.2023). 2023. URL: <https://datascientest.com/de/kuenstliche-intelligenz>.
- [4] W3Schools. *Machine Learning - Train/Test*. Einsichtnahme: (11.01.2023). 2023. URL: https://www.w3schools.com/python/python_ml_train_test.asp.