Produktbeschreibung:

**Motivation**

Der Trend der Technik geht immer mehr und mehr in Richtung Automatisierungstechnik. Neuste Entwicklungen zeigen großes Potential der Künstlichen Intelligenz. Dieses Potential möchte das ARS Electronika Center seinen Besuchern näherbringen.

Der derzeitige Stand der Technik ermöglicht es Menschen mithilfe von Text herausragende Bilder zu generieren. Auch ist es möglich kreative Texte zu schreiben. Was den Algorithmen jedoch fehlt ist die Eigenschaft an Eigenkreativität. DrAI löst das Problem durch eine Kombination aus verschiedenen Deep Learning Netzwerken kombiniert mit simplen Algorithmen. Da Menschen Materielle Lebewesen sind findet das ganze interaktiv durch unseren Roboter statt, welcher die generierten Bilder auf ein Blatt Papier zeichnet.

**Ideenfindung**

An der HTL-Neufelden wird eine Abschlussarbeit benötigt, um die Schule zu absolvieren. Samuel Nösslböck brannte schon seit er ein kleiner Junge war für Roboterbau und die Programmierung dafür. Rene Schwarz zeigte schon seit Jahren Interesse im Bereich Künstliche Intelligenz. Beide haben wir schon diverse Projekte in diesen Bereichen umgesetzt. Dies erbaute uns einen Grundwissensstock, welcher uns bei diesem Projekt weiterhilft. Da wir beide Freunde und Schulkollegen sind waren wir kurzerhand entschlossen die Abschlussarbeit gemeinsam umzusetzen. An unserer Schule wird immer eine Firma als Projektauftraggeber benötigt und so beschlossen wir das ARS Electronica Center und die Universität Linz zu fragen, ob sie ein Projekt für uns hätten.

Unserer Betreuungslehrer, Peter Rachinger, hat einen Kontakt zum ARS Electronica Center und so bekamen wir ziemlich schnell einen Ansprechpartner. Da sie kein konkretes Projekt hatten bekamen wird die Aufgabe, die wie folgt lautete:

„Wir sollten ein Projekt umsetzen, welches den Besucher Künstliche Intelligenz näherbringt.“

Wir überlegten uns diverse kreative Ideen und besprachen, diese mit unserem Ansprechpartner beim Ars Electronica Center und unseren Projektbetreuer Rachinger Peter. Da keine der Ideen alle Ansprüche und Wünsche von allen Personen erfüllte, brainstormten wir von selbstkomponierenden Systemen über Gestikenimmitierende Marionetten bis Samuel auf die Idee kam etwas mit Zeichnungen zu versuchen. Nach einigen weiteren Überlegungen entstand die finale Idee, ein von Künstlicher Intelligenz angetriebener kollaborativer zeichnender Roboter.

Ein Bild, das Entwurf, Zeichnung, Diagramm, technische Zeichnung enthält.

Automatisch generierte Beschreibung

Erste Konzeptskizze

Der Prozess zur finalen Zeichnung erfolgt wie folgt:

1. Der Besucher zeichnet eine Zeichnung mit einem Stift auf ein Blatt Papier
2. Der Roboter erstellt eine digitale Version davon
3. Die Künstliche Intelligenz erkennt was gezeichnet wurde und überlegt sich was hinzugefügt werden könnte
4. Linien werden aus der Zeichnung extrahiert
5. Der Roboter zeichnet das Überlegt auf ein Blatt Papier

Dies führt dazu, dass der Besucher in den Prozess integriert wird und mit der Künstlichen Intelligenz gemeinsam eine Zeichnung entwickelt.

Recherche

Nach einer Internett Recherche fanden wir heraus, dass es zwar schon Zeichenroboter gibt, die auf gewisse Datensätze trainiert wurden, jedoch funktionieren diese nur auf kleine Datensätze / Zeichnungen und wiesen nicht die von uns gewünschte Vielfältigkeit und Kreativität auf. Auch existierten in den meisten Fällen nur eine Software und keine Roboter, welcher die Zeichnung auf ein Blatt Papier bringt.

Zeitplan:

Ein Bild, das Text, Screenshot enthält.

Automatisch generierte Beschreibung

Neuheitsgrad

Die Idee des Projektes ist es zu einer bestehenden Zeichnung etwas Kreatives hinzuzufügen und den Menschen mit der KI kollaborieren zu lassen. Das Projekt sollte zur Schaustellung von den Möglichkeiten von Künstlichen Intelligenz und Mensch dienen. Zu Beginn des Projektes waren wir der Meinung künstliche Intelligenz wird in den nächsten Jahren zu einem neuen Werkzeug für die Menschheit, wie die Erfindung des Taschenrechners. Wie im Absatz Ideenfindung beschrieben gibt es diverse Algorithmen, mit denen sich durch Menschliche Kreativität Herausragendes entwickeln lässt. Durch Programme wie Github Copilot lassen sich komplexe Programme in beliebigen Programmiersprachen durch eine Textbeschreibung schreiben. Oft genügt eine einfache textliche Beschreibung des Lösungsansatzes und schon liefert ein Large Language Model wie GPT-4 den Programmcode. Was jedoch hierbei wichtig ist, ist die Beschreibung des Lösungsansatzes. Neuronale Netzwerke sind nur so gut wie ihre Trainingsdaten es sind. Befindet sich das Problem nicht in den Datensätzen so fällt es dem Programm meist schwer eine komplexe Lösung zu finden, die auch wirklich funktioniert. Hierbei wird dann die Menschliche Kreativität benötigt, um einen Lösungsansatz zu finden, der dann beschrieben wird und in Code umgesetzt wird. Mit Bildern ist es das gleiche, es wird Menschliche Kreativität benötigt, um kreative Textbeschreibungen für Bilder zu finden, die dann gezeichnet werden. Je weiter wir unseren Algorithmus entwickelten, desto mehr sahen wir, dass diese These nicht wahr ist. Unser Programmcode hat einerseits die Möglichkeit aus bestehendem etwas Neues zu erzeugen und andererseits aus einem weißen Blatt Papier ohne Textliche Beschreibungen und Menschliches einwirken kreative Bilder zu malen. Es ist eine ethische Frage ob Maschinen Kreativ sein können oder nicht, jedoch sehen die Resultate sehr vielversprechend aus. Unser Projekt beschränkt sich nur auf Bilder jedoch öffnet es die Frage wie lange es noch dauern wird, bis Maschinen eigene Lösungsansätze finden.

**Derzeitiger Stand:**

**Roboter:**

Derzeit befindet sich der Roboter in der Bau-Phase. Der Roboter setzt sich aus einem Grundrahmen aus Aluminium zusammen und wird mit 3D-Druckteilen verkleidet. Der Grundrahmen wurde schon gefertigt und ist fertig zusammengeschraubt. Die Abdeckungen sind noch in der Fertigung.

Ein Bild, das Entwurf, Design enthält.

Automatisch generierte Beschreibung

Fertiggestellter Roboter

**KI:**

**Version 1, Eigenes Neuronales Netzwerk**

Da in unserem Anwendungsfall keine Flächen mit Farbe bemalt werden sollten, kam die Idee auf, ein eigenes Generativ Adversal Network zu trainieren. Dies ist eines der ersten vielversprechenden Konzepte im Deep Learning Bereich. Hierbei werden zwei Netzwerke trainiert. Der Generator und der Discriminator. Der Generator erzeugt aus einem Random Noise oder einem Eingangsbild ein modifiziertes Ausgangsbild. Der Discriminator erhält als Eingangsparameter ein echtes oder ein generiertes Bild und wird darauf trainiert zu erkennen, ob das Eingangsbild echt oder generiert worden ist. Im besten Fall lernt der Generator so gute Bilder zu erzeugen, dass der Discriminator nicht mehr zwischen echt und generiert unterscheiden kann. Ein Bild, das Text, Screenshot, Diagramm, Säugetier enthält.

Automatisch generierte Beschreibung

Abbildung 5.3.1 GAN-Architektur

Da wir nur beschränkte Rechenleistung besitzen, haben wir uns zu Beginn darauf geeinigt, nur Landschaftsbilder als Trainingsdaten zu verwenden. Da der Benutzer nur Umrisse zeichnet, wurden verschiedene Kantenerkennung Algorithmen verglichen.

Da der Benutzer jedoch keine vollständigen Bilder zeichnet und es die Aufgabe des DrAI-Roboters ist Konturen zu der Originalzeichnung hinzuzufügen, wurden Bildbereiche / Linien aus dem Originalbild entfernt. Der Generator sollte lernen, aus dem bearbeiteten Bild das Originalbild zu rekonstruieren. Da der Generator etwas erzeugen sollte, werden zufällig normalverteilte Werte dem Bild hinzugefügt.

Ein Bild, das Entwurf, Zeichnung, Lineart, Kunst enthält.

Automatisch generierte BeschreibungEin Bild, das Entwurf, Lineart, Zeichnung, Clipart enthält.

Automatisch generierte Beschreibung

Eingangsbild (Input) Originalbild (Output)

Die Vorteile des Algorithmuses wären:

Wenig Rechenaufwendig während der Ausführung

Auf den Kundenwunsch hin zuschneidbar

Auf unsere Anwendung anpassbar

Die Nachteile des Algorithmuses wären:

Rechenaufwendig zu Trainieren

Aufwendig Daten zu bekommen

Nicht sehr kreative Bilder

Es wären nur zum Beispiel Landschaftsbilder möglich

Schnell erstellte Resultate waren wenig vielversprechend. Es war kein Hauch von Kreativität zu erkennen und der Algorithmus drohte einfach nur schwarze Bilder zu generieren, da der Discriminator zu gut erkannte, welches Bild echt und welches generiert wurde. Dies führte dazu, dass der Fehlerwert des Generators weiter stieg und so die Bilder nicht optimiert wurden.

**Version 2, Stable Diffusion**

Um die Probleme, die ein selbst trainierter Algorithmus liefert, zu lösen, wurde Stable Diffusion benutzt. Dies ist ein Diffusion-Algorithmus vor trainiert von Stability Ai. Dieser Algorithmus benutzt Text, um Bilder zu erstellen. Hierbei wird wie bei einem Generativ Adversarial Network aus einem zufällig normalverteilten Bild ein für uns Menschliches Bild erzeugt. Der Algorithmus wird trainiert indem zu einem für das menschliche Auge Sinnhaft Bild normalverteilte Werte hinzuaddiert werden. Diese Werte werden immer höher und das Bild verliert mit der Zeit die Sinnhaftigkeit. Ziel des Netzwerkes ist es nun über den Verlauf der Zeit den Diffusion Process umzukehren und basierend auf dem diffundierten Bild und dem Text Prompt das Originalbild wieder herzustellen.

Ein Bild, das Text, Screenshot enthält.

Automatisch generierte Beschreibung

Ein Bild, das Text, Screenshot, Schrift, Diagramm enthält.

Automatisch generierte Beschreibung

Die erste Version dieses Netzwerkes wurde auf den Laion2B-en Datensatz trainiert. Dies ermöglichte es dem Netzwerk eine Vielzahl an Bildern zu konstruieren und viele Bereiche, wie zum Beispiel Landschaften, Städte, … zu rekonstruieren, basierend auf Textbeschreibungen. Je kreativer die Textbeschreibung, desto kreativer das generierte Bild.

Hierbei tritt das Problem auf, das der Algorithmus hervorragende Bilder erzeugen kann, jedoch hängt die Kreativität von der Textbeschreibung ab.

Version 3, InterStableCloud

Ein Bild, das Entwurf, Zeichnung, Fahrzeug, Clipart enthält.

Automatisch generierte Beschreibung

Ein Bild, das Entwurf, Zeichnung, Fahrzeug, Lineart enthält.

Automatisch generierte Beschreibung

Konstruktion

AI

Innovatives:

Lösungsweg

Einzigartigkeit

Neuheitsgrad

Anwendungsgebiet

Erkenntnisgewinn, Anwendungsgebiete, soziale Auswirkungen

Herkunft der Idee

Kommerzielle Verwertbarkeit, Markttauglichkeit

Projekt Kompetenz