

Einleitung

Die Studienarbeit beschäftigt sich mit dem Thema Robotersteuerung durch Gestenerkennung mithilfe einer künstlichen Intelligenz. Grundlage für die Steuerung des Roboters ist eine LEGO Mindstorms, mit dessen Hilfe die Motoren angesteuert werden können. Die künstliche Intelligenz soll mithilfe eines Raspberry Pi betrieben werden, welche Signale mit dem LEGO Mindstorms austauschen kann.

Über eine Kamera sollen Bilder von einer Person aufgenommen werden, welche Buchstaben in Gebärdensprache vorführt. Die KI erkennt die gezeigten Buchstaben und übergibt das buchstabierte Wort an einen Lego Mindstorms Roboter, welcher das Wort auf einem Blatt Papier zeichnet. Beispielsweise kann die Person das Wort „Herz“ in Gebärdensprache buchstabieren, wodurch der Roboter ein Herz zeichnet. Eingeschränkt wird die Arbeit dadurch, dass zu Beginn nur einige einfache festgelegte Formen wie Kreis, Viereck etc. unterstützt werden.

Roboter

Der Roboter wird über ein Lego NXT gesteuert, an welchen Sensoren und Motoren angeschlossen werden können. Die Motoren sind für die Bewegungen des Stiftes zuständig. Ein Raspberry Pi 4 stellt die Plattform für die KI zur Verfügung.



Abb: Raspberry Pi Kamera

Mithilfe eines Kamera-Modules des Raspberry Pi können die Bilder direkt ohne Verzögerung beim Übertragen aufgenommen und für die KI bereitgestellt werden. Eine Verbindung zwischen Raspberry Pi und NXT ermöglicht, dass der Raspberry Pi den NXT steuern und die Anweisungen geben kann, wie er den Stift zu bewegen hat

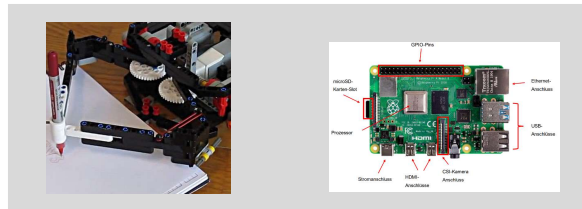


Abb: Lego Roboter

Abb: Raspberry Pi

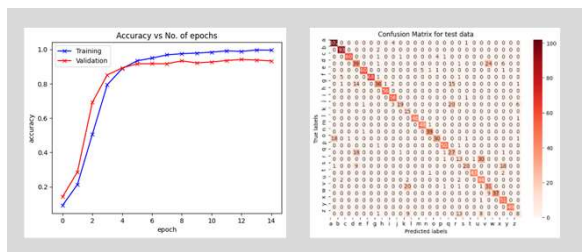
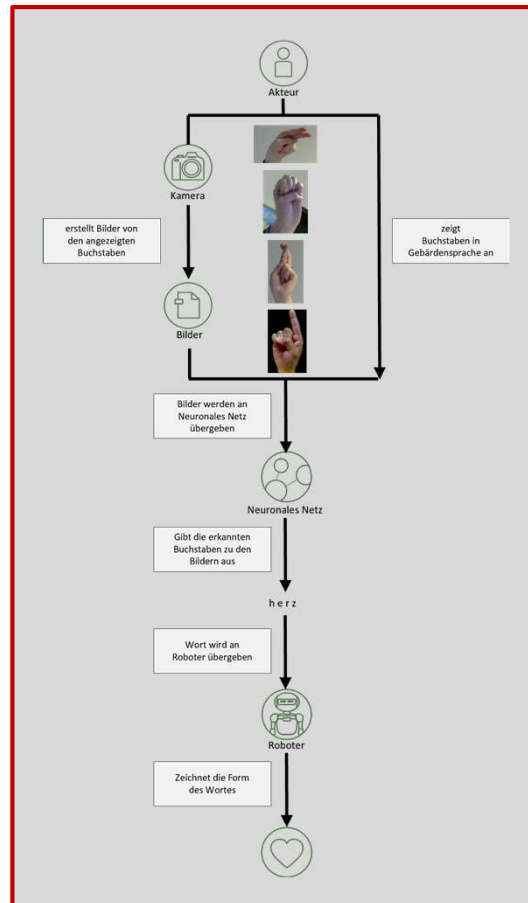


Abb: Verlauf der Accuracy

Abb: Confusion Matrix

KI-Modell

Zur Erkennung der Buchstaben in Gebärdensprache, wird ein Convolutional Neural Network (CNN) implementiert. CNNs sind eine spezielle Art von künstlichen neuronalen Netzwerken, die besonders für die Verarbeitung von Bildern und visuellen Daten optimiert sind. CNNs wurden entwickelt, um die Herausforderungen der Bilderkennung effizienter zu bewältigen, indem sie bestimmte Eigenschaften und Muster in Bildern erkennen können.

Die dafür benötigten Trainingsdaten stammen zum Teil aus einem öffentlichen Datensatz, der größte Teil wird jedoch aus der Aufnahme eigener Bilder gewonnen, die zudem durch verschiedene Augmentationen angereichert werden.

Um für das CNN bestmögliche Trainingsdaten zu generieren, werden die aufgenommenen Originalbilder auf die erkannte Hand im Bild zugeschnitten.



Abb: Ablauf zur Erstellung eigener Daten

Die Modellierung des CNN wird mit PyTorch umgesetzt. Es handelt sich dabei um ein sequentielles Modell. Als alternative Möglichkeiten wird die Verwendung vortrainierter Modelle wie Resnet und VGG19 in Betracht gezogen.

Das trainierte neuronale Netzwerk zur Erkennung von Buchstaben in Gebärdensprache erzielt eine Genauigkeit von 79% bei der Erkennung der Testdaten. Dies zeigt, dass das Modell erfolgreich ist und eine zufriedenstellende Leistungsfähigkeit bei der Identifizierung von Buchstaben, die in Gebärdensprache angezeigt werden, aufweist.

Kontakt

Projekt wird betreut von:

Peter Firmkäs

Duale Hochschule Baden-Württemberg

Ira Göppinger
Kim Junker
Benjamin Tomazic