Abschlussbericht für das Abschlussprojekt Deep Learning

Tobias Giesler, Florian Graf, Tim Kleinoth

TH Köln

Cologne, Germany

tobias\_ralf.giesler@smail.th-koeln.de

florian.graf@smail.th-koeln.de

tim.kleinoth@smail.th-koeln.de

*Zusammenfassung*— Dieser Bericht beschreibt das Erstellen einer Deep Learning Architektur, welche die Handzeichen „Stein, Schere und Papier“ erkennen soll. Verwendet werden dabei die Datensätze von „Julien de la Bruère-Terreault“ [1], sowie ein eigens erstellter Datensatz.

Schlagwörter—Deep Learning, Bilderkennung, Künstliche Intelligenz

# Einleitung

Der Begriff „Künstliche Intelligenz“ (KI) beschreibt Computerprogramme, die versuchen eine menschliche Intelligenz nachzuahmen. Ziel dabei ist es ein Programm zu entwickeln, dass selbstständig Probleme löst, indem es auf unterschiedlichste äußere Einflüsse reagiert. Der Begriff Deep Learning beschreibt dabei eine gewisse Form einer künstlichen Intelligenz, welche aus mehreren sogenannten „Hidden Layers“ besteht. Diese Schichten beschreiben die Vernetzung mehrerer „Perceptrons“ also künstliche Neuronen (siehe Abbildung 1), die zum Anlernen eines Prozesses verwendet werden.

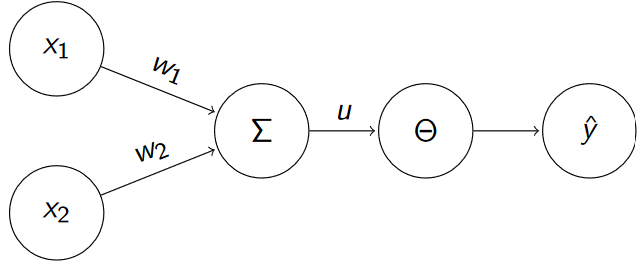


Abbildung 1: Darstellung eines Perceptrons [2]

Künstliche Intelligenzen finden schon heute in vielen Einsatzgebieten Verwendung. Zu diesen Einsatzgebieten zäht die Spracherkennung, Videoempfehlungen, Suchanfragenempfehlungen, oder (in der Industrie verwendet) „condition monitoring“, welches eine Überwachung von Sensorwerten durch eine künstliche Intelligenz beschreibt, um Ausfälle von Maschinen schon frühzeitig zu erkennen [3].

Ein weiteres großes Feld in dem künstliche Intelligenzen verwendet werden ist der Bereich der Bilderkennung. Bilderkennung beherbergt ein enormes Potential für Entwicklung. Bilderkennung kann und wird bereits in Bereichen wie Gesichtserkennung, autonomes Fahren, sowie die Überwachung von industriellen Anlagen verwendet.

In diesem Paper wird die Erstellung und Optimierung eines Deep Learning Algorithmus zum Erkennen von den Handzeichen „Stein, Schere und Papier“ beschrieben. Dabei wird auf unterschiedliche Möglichkeiten zur Verbesserung der sogenannten „Generalisierungsfähigkeit“ der KI eingegangen und deren Auswirkungen auf die Lernrate thematisiert.

# Aufbau des deep learning algorithmus

In diesem Kapitel wird der Grundaufbau des Algorithmus sowie seine Optimierungsmöglichkeiten zur Verbesserung der „Generalisierungsfähigkeit“ thematisiert. Die Generalisierungsfähigkeit bezeichnet dabei die Fähigkeit eines neuronalen Netzes das durch den Trainingsdatensatz Erlernte auf ein „fremden Datensatz“ (zum Beispiel dem Validierungsdatensatz) anzuwenden. Eine gute Generalisierungsfähigkeit ist damit ein Indikator für ein gut trainiertes neuronales Netz.

Des Weiteren wird auf den Trainingsdatensatz eingegangen. Grundlagen zum Aufbau noch hinzufügen?

## Aufbau und Ergebnisse der Lernrate einer nicht optimierten KI

Wie oben bereits erwähnt, soll Ziel dieses Projektes sein, eine Deep Learning Architektur zu entwerfen die selbstständig die Handzeichen „Stein, Schere, Papier“ erkennt. Dazu wurde eine Datensatz erstellt mit Bildern von den Handzeichen „Stein, Schere, Papier“ sowie „Rest“ welches ein den anderen drei Kategorien nicht zugehöriges Handzeichen darstellt. Zusätzlich wurde der bereits verwendete Datensatz von „Julien de la Bruère-Terreault“ verwendet. Ein Beispiel eines Fotos aus diesem Datensatz ist in Abbildung 2 dargestellt:



Abbildung 2: Beispielbild aus dem verwendeten Datensatz (Papier)

Zunächst wurde eine nicht optimierte Deep Learning Architektur verwendet. Diese (Aufbau der nicht optimierten KI beschreiben). Der beschriebene Algorithmus wurde dann mithilfe des Trainingsdatensatzes trainiert, wobei die Trainingsergebnisse mithilfe eines Validierungsdatensatzes zusätzlich überprüft wurden. Anhand der erzielten Ergebnisse der entworfenen Deep Learning Architektur im Trainingsdatensatz sowie im Validierungsdatensatz kann man gut ein gewisses „Lernverhalten“ der entworfenen Architektur begutachten und beurteilen. Die Ergebnisse, die mit dem nicht optimierten Datensatz erzielt wurden, werden in Abbildung 3 und Abbildung 4 dargestellt:

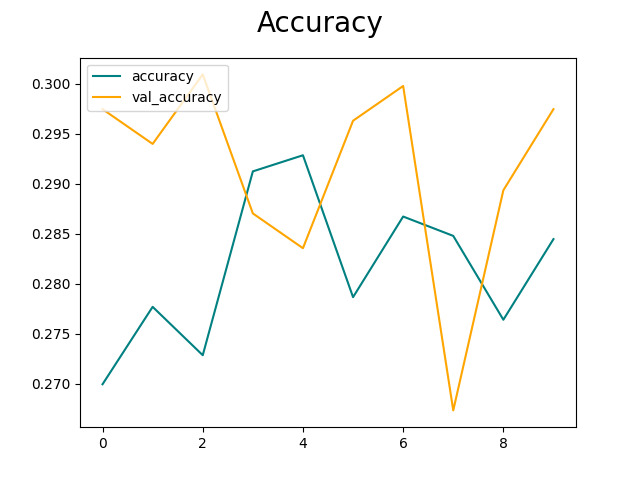


Abbildung 3: Testergebnis der Genauigkeit der nicht optimierten Deep Learning Architektur (Achsen noch beschriften! Genauere Legende, Deutscher Achsentitel)

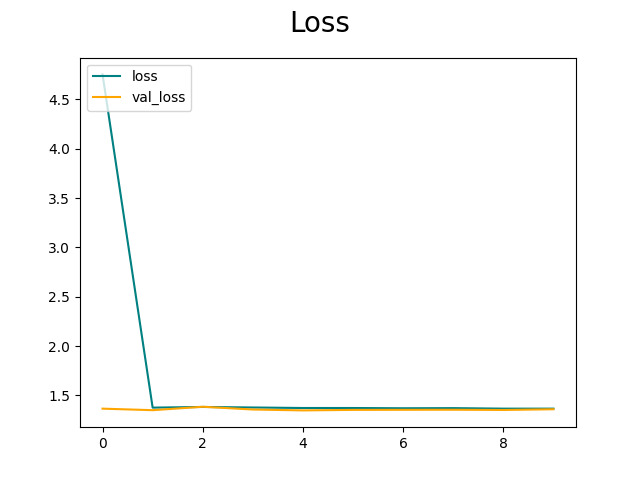


Abbildung 4: Testergebnisse der Loss Funktion der nicht optimierten Deep Learning Architektur (Besser andere Ergebnisse, da loss Funktion der Validation nicht funktioniert)

Interpretation der Ergebnisse hinzufügen.

## Early Stopping

Eine der Verbesserungsmöglichkeiten der Generalisierungsfähigkeit ist das sogenannte „Early Stopping“. Early Stopping wird verwendet, um zu vermeiden, dass sich das entworfene neuronale Netz zu sehr an einen Trainingsdatensatz anpasst und damit übertrainiert wird (overfitting). Ein übertrainiertes Netz erzielt zwar im Trainingsdatensatz, durch die Überanpassung, bessere Ergebnisse, erzielt aber im bisher fremden Validierungsdatensatz schlechtere Ergebnisse. Das trainierte Netz hat somit eine schlechtere Generalisierungsfähigkeit. Early Stopping kann hierbei gut ansetzen. Beim Early Stopping wird die Fehlerfunktion des Validierungsdatensatzes überwacht. Wird bei der Überwachung über einen zuvor definierten Zeitraum eine signifikante Verschlechterung der Fehlerfunktion des Validierungsdatensatzes festgestellt, während die Fehlerfunktion des Trainingsdatensatzes sich verbessert, so wird das Training beendet und der Algorithmus gibt anschließend die Parameter zurück, die vor dem Zeitpunkt der Verschlechterung verwendet wurden.

##### Quellenangaben

1. Datensatz von Julien de la Bruère-Terreault: <https://www.kaggle.com/datasets/drgfreeman/rockpaperscissors>
2. Jan Salmen, Vorlesungsfolien „Deep Learning” S. 31
3. Geron Aurelien, “Hands-on Machine Learning with Scikit-Learn, Keras & TensorFlow” S. XV