Technology
Arts Sciences
TH Köln

Technische Hochschule Köln
Fakultät für Informatik und Ingenieurwissenschaften

Praxisprojekt WS18/19 - Dokumentation

Entwicklung einer Objekterkennung für Regal-Systeme, mit Hilfe von Bildverarbeitung, in einem Neuronalen Netz

Vorgelegt an der TH Köln
Campus Gummersbach
im Studiengang
Medieninformatik

ausgearbeitet von:
Torben Krause
(Matrikelnummer: 11106885)

Prüfer/Betreuer: Prof. Dr. Martin Eisemann

Gummersbach, 28.12.2018

Inhaltsverzeichnis

Abbildungsverzeichnis

1 Einleitung

Im Rahmen des Praxisprojektes in der Medieninformatik, geht es um die Definition und Umsetzung einer Projektidee, mit Hilfe zuvor erlernter Grundlagen aus den Modulen des Studiums und der Erarbeitung neuer Themen. Folgende Dokumentation beschreibt die ersten Entwicklungsschritte von der Ideenfindung über den Konzeptentwurf bis hin zur Durchführung. Sie umfasst neben einer Zieldefinition, die ersten Ergebnisse der Entwicklungsphase, sowie einer Evaluation der Trainingsphase im Hinblick auf die Anforderungen. Die bis zu diesem Projektstand ausgearbeiteten Ergebnisse sollen den Projektverlauf mit den getroffenen Abwägungen und Alternativen verdeutlichen. Zur Ausführung des Konzepts wird auf ein Projektplan gesetzt, der geplante Aktivitäten festlegt und gewünschte Ergebnisse in einem zeitlichen Rahmen definiert.

1.1 Problemstellung

Das richtige Verwalten von Lebensmitteln, welche gekühlt im Kühlschrank aufbewahrt werden müssen, ist nicht immer leicht. Oft müssen schlecht gewordene Lebensmittel weggeschmissen werden, weil sie vergessen werden oder es fehlen Lebensmittel, welche gebraucht werden. Gerade in Haushalten mit mehreren Personen, wie zum Beispiel Familien oder Wohngemeinschaften, ist das Kühlschrankmanagement schwierig, da nicht jeder Bewohner weiß welche Lebensmittel aus dem Kühlschrank genommen wurden und welche hinein gelegt werden. Die Kommunikation dafür fehlt in den meisten fällen. Aus diesem Grund werden schnell verderbende Lebensmittel leicht vergessen oder Produkte durch fehlerhafte Absprache mehrfach oder nicht gekauft.

1.2 Zielsetzung

Im Laufe des Praxisprojektes soll ein Künstliches neuronales Netz entwickelt werden, welches den Inhalt eines Kühlschranks mit Hilfe einer Kameraaufnahme und anschließender Bilderkennung bestimmen kann. Dabei soll ermittelt werden, welche Lebensmittel sich im Kühlschrank befinden und seit wann diese gelagert werden. Außerdem könnten diese Informationen die Möglichkeit für weitere Anwendungen dienen.

1.2.1 Geplante Funktionsweise

Der Kühlschrankinhalt soll mithilfe einer drahtlosen Kamera, welche im inneren des Kühlschranks angebracht ist, aufgezeichnet werden. Beim schließen der Kühlschranktür soll unter Verwendung eines Kontaktsensors ein Bild aufgezeichnet und an ein externes System gesendet werden. Um ein gut sichtbares Bild zu erhalten, sollen verschiedene Möglichkeiten zur Verbesserung der Lichtverhältnisse evaluiert werden. Für die Bilderkennung so wie einer Klassifizierung sollen verschiedene Ansätze verglichen und anschließend bewertet

werden. Durch die Klassifizierung soll das System alle Produkte, welche sich auf dem Bild befinden identifizieren und einordnen können. Aus den Ergebnissen der Analyse wird nun eine Produktliste vom Inneren des Kühlschranks erstellt und anschließend in einer Datenbank gespeichert. Das User Interface soll als mobile Anwendung implementiert werden, auf der alle erhobenen Informationen dem Benutzer in Form einer Liste zur Verfügung gestellt werden. Dabei kann der Benutzer aktuelle Informationen vom Server abrufen.

Die ersten Schritte um das geplante System zu modellieren, ist die Entwicklung eines künstlichen neuronalen Netzes. Aus diesem Grund wird der Fokus für das Praxisprojekt, auf das künstliche neuronale Netz gelegt. Ziel des Projektes ist die Entwicklung eines neuronalen Netzes welches mehrere vordefinierte Klassen zuverlässig erkennt und zuordnen kann. Das Netz soll außerdem so konzipiert werden, das Produkte unabhängig ihrer Position im Bild erkannt werden können. Außerdem sollte das Neuronale Netz auch mit leicht verdeckten Objekten zurecht kommen. Ein Optionales Ziel welches erstrebenswert wäre ist eine Unempfindlichkeit gegenüber Licht- und Bildqualitätsverhältnissen.

1.3 Anforderungen

In diesem Kapitel, sollen die ersten Anforderungen an das zu entwickelnde künstliche neuronales Netz, welche mit Hilfe der Anforderungsschablonen von 'Die SOPHISTen'[QUELLE] verfasst wurden. Diese Anforderungen sind nicht endgültig oder gar vollzählig, da dass das Ziel für dieses Projekt kein voll funktionsfähiger Prototyp sein soll und weiter Anforderungen sich aus der Weiterentwicklung und Vervollständigung des Systems ergeben.

1.3.1 Funktionale Anforderungen

- Das System muss fähig sein, eingegebene Informationen in Bild- und Videoformat aufzunehmen und zu verarbeiten.
- Das System muss fähig sein, die Trainierten Klassen zu erkennen und einzuordnen.
- Das System soll fähig sein, mehrere Objekte auf einem Bild gleichzeitig erkennen und und deren Position festzustellen.

1.3.2 Organisationale Anforderungen

- Das System wird so gestaltet sein das die Existierenden Klassen unkompliziert erweitert werden können.
- Das System wird so gestaltet sein, das es mit wenig aufwand betrieben werden kann.
- Die Installation benötigter Module, soll für den Benutzer so einfach wie möglich gemacht werden.
- Das System soll so gestaltet sein, das es einfach auf ein Endgerät transferiert werden kann.

1.3.3 Qualitative Anforderungen

- Das System soll so ausgewählt werden, das die Geschwindigkeit der Objekterkennung möglichst hoch ist.
- Das zu Trainierende Netz, muss eine möglichst hohe Genauigkeit aufweisen.
- Das zu entwickelnde Netz, soll auf den Aktuellen Plattformen arbeiten und neue Technologien unterstützen.
- Das system soll fähig sein, auch leicht verdeckte Objekte richtig zu identifizieren.
- Das System soll fähig sein, auch bei schlechten Bildverhältnissen gute Ergebnisse zu liefern.

2 Künstliche Neuronale Netze

Der Begriff Neuronale Netze kommt aus der Biologie und beschreibt die Funktionsweise des Menschlichen Gehirn. Das Gehirn eines Menschen besteht aus ungefähr 10^{11} Nervenzellen[QUELLE], den so genannten Neuronen, welche untereinander verbunden sind. Diese nehmen Informationen auf. Wenn deren Schwellwert, welcher bei jedem Neuron unterschiedlich ist, überschritten wird, gibt er einen Wert an das nächste Neuron weiter. Das Netz ist unter anderem für das menschliche Lernen zuständig. Das bedeutet, dass ein neuronales Netz eine Verbindung aus vielen Nervenzellen bildet, welche erhaltene Informationen verarbeiten und weiter leiten.

Ein künstliches neuronales Netz ist dementsprechend eine Nachbildung eines Menschlichen Gehirns welches auf Grundlage mathematischer Formeln arbeitet. Auf die verwendeten Formeln, wird in diesem Kapitel nicht eingegangen, eher soll die Funktionsweise erläutert werden. Künstliche Neuronale Netze (KNN) stellen einen Bestandteil des Maschinellen Lernens dar. Oft werden KNN,s für die Klassifizierung von Datensetzen (Bildern, Statistiken, Profile in Sozialen Netzwerken) verwendet. Netze welche auf eine Reihe von Klassen trainiert wurden, liefert nach Eingabeinformationen, eine oder mehrere Wahrscheinlichkeiten als Ausgabe. Diese Ausgabe soll dabei bestimmen, wie hoch die Wahrscheinlichkeit ist, das es sich bei der Eingabe um die Ermittelte Klasse handelt. Als Eingabe kann es sich Beispielsweise um ein Bild oder Video handeln. Die Klassen auf welche das Neuronale Netz Trainiert ist, können dabei Hunde und Katzen sein. Das Künstliche Neuronale Netz soll also berechnen ob es sich auf dem eingegebenen Bild um eine Katze oder einen Hund handelt.

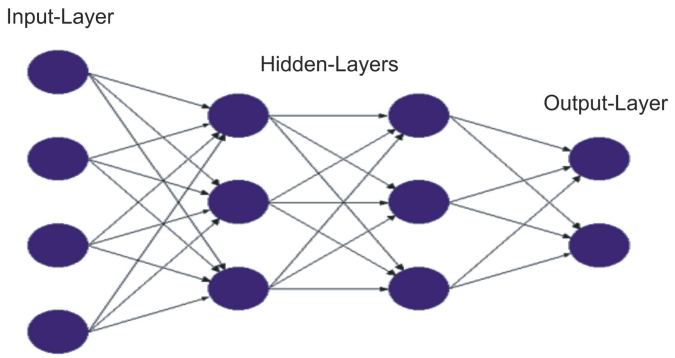

Abbildung 1: Schematischer Aufbau eines Künstlichen neuronalen Netzes

Der Aufbau eines Künstlichen Neuronalen Netzes kann grundsätzlich in drei Schichten aufgeteilt werden (Abbildung 1). Die erste Schicht, ist der Input Layer. Dieser Layer bekommt

alle Eingangsdaten von einem Bild. Jedes Pixel auf dem Bild wird mit einem Neuron verbunden, dabei hat jedes Neuron eine andere Gewichtung und Schwellwert bei dem er Informationen weiter leitet. Wird der Schwellwert erreicht, gibt das Neuron die verarbeiteten Informationen an die nächste Schicht weiter. Die zweite Schicht besteht aus einem oder mehreren Hidden Layer. Jeder Layer ist in dieser Schicht auf verschiedene Merkmale trainiert. Anfangs werden nur grobe Strukturen erkannt, wie Beispielsweise Geraden oder Kanten. Diese Strukturen werden detaillierter, desto mehr Layer durchlaufen wurden. Die dritte Schicht besteht aus dem Output Layer. Die Muster die in den vorherigen Layern erkannt und zusammengesetzt wurden, vergleicht der Output Layer mit den trainierten Klassen und gibt eine prozentuale Wahrscheinlichkeit aus, um welche Klasse es sich in dem Bild handelt.

2.1 Convolutional Neural Network

Anders als bei einem KNN, welches auf Grundlage des Menschlichen Gehirns entworfen wurde, orientiert sich ein Convolutional Neural Network (CNN) an dem Konzept des menschlichen Sehens. Dabei werden maschinell, kleine Bereiche visueller Informationen so simuliert, als wären sie benachbarte Sehnerven im Auge[QUELLE]. Gerade für die Verarbeitung von Bilddaten werden häufig Convolutional Neural Networks eingesetzt. Voraussetzung für ein CNN ist die Darstellung und Anordnung der Datenpunkte. Diese müssen eine matrixartige Anordnung von Werten haben, wie beispielsweise Pixel.

Bei einem Convolutional Layer (Abbildung 2), oder auch im Deutschen Faltungsschicht, werden über die Pixel des Eingabebildes eine Schablone gelegt. Diese wird je nach Bildgröße, mehrere male Horizontal wie auch Vertikal verschoben. Diese Layer haben meist eine ungerade quadratische Abmessung (3x3, 5x5, 7x7). Die frühen schichten bilden zunächst ein Skalarprodukt der betroffenen Pixel und können dadurch grobe kanten erkennen. Jeder Layer ist dabei auf eine bestimmte Art von Mustern trainiert. Je mehr Layer durchlaufen, desto genauer werden die Muster welche das neuronale Netz erkennt und umso mehr wird das Eingabebild runter skaliert.

Der Pooling Layer (Abbildung 2) funktioniert ähnlich wie der Convolutional Layer. Auch bei dieser Methode wird eine Schablone über die Pixel des Bildes gelegt und Horizontal wie auch Vertikal verschoben. Hierbei werden die Farbwerte der beteiligten Pixel verarbeitet. Es kann entweder der höchste Farbwert eines Pixels übernommen, oder der durchschnittliche Farbwert berechnet werden. Die letzten Schichten des Netzes bestehen aus Fully Connected Layer welche den normalen Aufbau eines künstlichen neuronalen Netzes darstellt. Das bedeutet, alle Ausgaben werden wie beim KNN mit jedem Neuron verbunden. Die letzte Schicht übergibt seine Ausgaben an den Output Layer, welcher die Wahrscheinlichkeiten der Objekte berechnet.

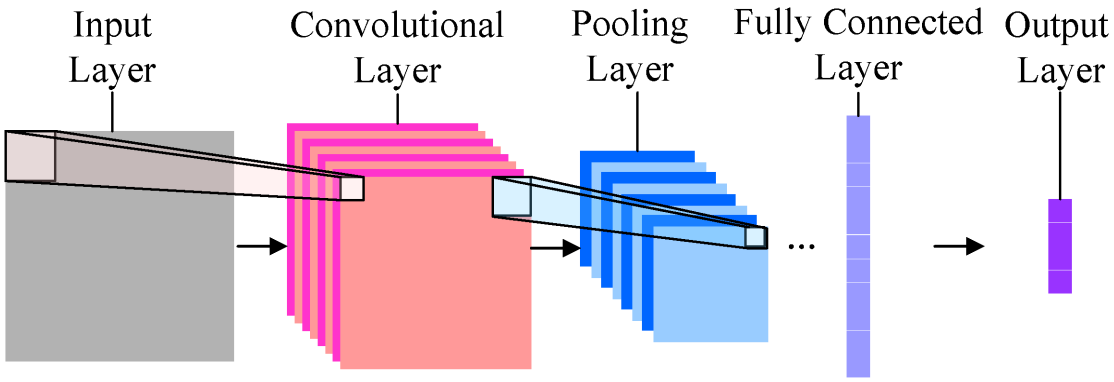

Abbildung 2: Funktionsweise eines Convolutional Neural Network

2.2 Abwägung von Neuronalen Netzen

Für das Entwickeln eines neuen neuronalen Netzes wird eine hohe Anzahl an Trainingsdaten sowie viel Rechenleistung benötigt. Bestehende Netze wurden mit mehreren 100 Tausend Bilden, Tage, bis Wochen lang trainiert. Eine andere Möglichkeit besteht darin, frei verfügbar trainierte Netze zu verwenden. Diese Netze haben für ihre bis zu 80 Klassen eine Genauigkeit von durchschnittlich 90 Prozent oder eine mAP von 30. Einige dieser trainierten Netze werden von Unternehmen als Open-Source zur Verfügung gestellt.

Um ein eigenes Netz mit einer annähernd hohen Genauigkeit zu erreichen ist in der Hinsicht auf die eigenen Ressourcen schwierig. Aus diesem Grund, soll für das Projekt ein bestehendes Neuronales ausgewählt werden welches mit eigenen Klassen und Trainingsdaten trainiert wird. Die Kriterien für das ausgewählte Netz, sind zum einen eine hohe Genauigkeit und eine moderate Geschwindigkeit haben mit der das spätere System effizient arbeiten kann. Der wichtige Aspekt ist die Genauigkeit und Verlässlichkeit der Prognosen. Weitere Anforderungen wurden in den vorherigen Kapiteln bereits aufgestellt.

In der folgenden Abwägung, wurden einige verfügbare Netze aufgeführt, die für Tensorflow zur Verfügung gestellt werden. Dabei werden die aufgeführten Modelle in Genauigkeit und Geschwindigkeit verglichen.

2.2.1 Auswahl des Netzes

In Hinsicht auf Trainierte Netze bietet T̈ensorflowëinige Möglichkeiten an. Hierbei handelt es sich um die T̈ensorflow detection model zoo[QUELLE] welche mit verschiedensten Datensätzen und Durchläufen trainiert wurden. Die Datensätze die verwendet wurden sind unter anderem das COCO dataset, das Kitti dataset, das Open Images dataset, das AVA v2.1 dataset und das iNaturalist Species Detection Dataset. Die COCO-Models bieten das größte Spektrum an Netzen. Bei dem COCO Datenset handelt es sich um 'Common

Objects In Context' und beinhaltet rund 330 Tausend Bilder, welche über 1.5 Millionen schichten in 90 Klassen trainiert wurden. Da es sich um häufige auftretende Objekte in einer realen Umgebung handelt, sind die Modelle gut für unser vorhaben geeignet. In der folgenden Tabelle, wurden ein paar dieser Modelle ausgesucht und verglichen.

Modell Name	Geschwindigkeit	COCO mAP	Ausgabe
faster_rcnn_inception_v2_coco	58 ms	28	Boxen
faster_rcnn_resnet50_coco	89 ms	30	Boxen
rfcn_resnet101_coco	92 ms	30	Boxen
faster_rcnn_resnet101_coco	106 ms	32	Boxen
ssd_mobilenet_v1_fpn_coco	53 ms	32	Boxen
ssd_resnet_50_fpn_coco	76 ms	35	Boxen

Tabelle 1: Trainierte Modelle auf dem COCO Dataset [QUELLE]

In der Tabelle, werden 6 mögliche Modelle aufgeführt. Beschrieben werden diese durch eine Geschwindigkeit in welcher das Neuronale Netz im durchschnitt arbeitet, einer Genauigkeit und eines Ausgabetyps. Die zweite Spalte, COCO mAP (Mean Average Precision) gibt die mittlere durchschnittliche Genauigkeit an, mit welcher es Objekte erkennt. Berechnet wird diese durch die Genauigkeit der richtigen Ergebnisse im Vergleich auf alle ausgeführten Durchläufe des Systems. In der Letzten spalte wird die Ausgabeform beschrieben. Hierbei kann zwischen Boxen und Masken unterschieden werden. Dabei wird je nach Ausgabetyp entweder ein viereckiger Kasten (Abbildung **??**) um das object erstellt, oder es wird eine Maske (Abbildung **??**) über dieses gezogen.

(a) Makierung durch Maskierung (b) Makierung durch Kästen

Abbildung 3: Ausgaben von Neuronalen Netzen

Bei dem Vergleich der Aufgezählten Modellen fällt auf, das alle eine ähnliche Genauigkeit aufweisen und sich leicht unterscheiden. In der Geschwindigkeit jedoch werden die Unterschiede erkennbar größer. Durch die hohe Geschwindigkeit und vergleichsweise hohe Genauigkeit, wird vorerst das faster_rcnn_inception_v2_coco Modell getestet, und ausgewertet.

Das ausgewählte Modell wurde für Testzwecke mit 6 Klassen um-trainiert. Die Klassen bestanden aus einem Skart-kartenset von Neun bis Ass [QUELLE]. Rund Fünf Stunden wurde das Netz trainiert, und hat dabei 60 tausend Testdurchläufe absolviert. Eine Überprüfung des Netzes hat eine Genauigkeit von 97 bis 99 Prozent ergeben. Auch Tests ergeben das die Erkennung fehlerfrei funktioniert. Das Testbild (Abbildung 4 zeigt, das alle Karten mit einer hohen Wahrscheinlichkeit identifiziert werden konnten.

Abbildung 4: Analysiertes Bild, durch das Testnetz

Das erfolgreiche Training und testen des Netzes hat die Verwendung des Ausgewählten Modells bestärkt. Die Vorgehensweise sieht in den weiteren Schritten eine Definition und Festlegung der Trainings-Daten für das benötigte neuronale Netz.

3 Durchführung

In diesem Kapitel der Dokumentation, wird die Praktische Durchführung beschrieben und Entscheidungen aufgeführt. Dabei werden die verwendeten Frameworks und die Entwicklungsumgebung aufgezeigt und erläutert. Im weiteren geht es um das festlegen der Klassen für das Netz und die anschließende Durchführung.

3.1 Entwicklungsumgebung

Um ein bestehendes neuronales Netz neu zu trainieren werden neben dem dem Neuronalen Netz weitere Frameworks sowie eine Entwicklungsumgebung benötigt. In den kommenden Unterpunkten wird die Entwicklungsumgebung sowie alle benötigten Bibliotheken aufgeführt.

3.1.1 Verwendete Hardware Konfiguration

Das Netz wurde auf einem Windows 8.1 Betriebssystem trainiert. Berechnet wurden die Trainingsschritte von einer Nvidia GeForce GTX 980 GPU mit 4GB VRAM, einer Intel Core i5 CPU und 16GB RAM (Arbeitsspeicher). Für einen Trainingsschritt, mit dem ausgewählten Netz, benötigt die Grafikkarte durchschnittlich 250ms. Im Vergleich hat die CPU 2 Sekunden pro Schritt benötigt. Die Grafikkarte arbeitet somit 8 mal schneller durch den Vorteil mit Matrixmultiplikationen und soll für das Training verwendet werden.

3.1.2 Anaconda python

Das verwendete CNN wurde in der Programmiersprache Python erstellt, weswegen wir zum um-trainieren mit Python arbeiten. Des weiteren bietet Python eine Vielzahl an verfügbaren Bibliotheken welche wir benötigen. Anaconda ist eine Öpen-Source-DistributionFür die Programmiersprache Python. Mit dieser können wir gleichzeitig mehrere Versionen von Python generieren und die notwendigen Bibliotheken installieren. Anaconda bietet den Vorteil einfach zwischen den verschieden Python Versionen wechseln zu können
In diesem Projekt wurde mit der Python version 3.6.7 gearbeitet.

- Installation von Python 3.6.7: conda create -n <Name> python=3.6.7

3.2 Frameworks

Um das Künstliche neuronale Netz zu trainieren, wird ein Framework und mehrere Programm-Bibliotheken benötigt. Die Auswahl wird im folgenden aufgezählt und kurz erläutert.

3.2.1 Tensorflow

Bei Tensorflow handelt es sich um eine Plattform unabhängige Programmbibliothek unter Open-Source-Lizenz, die sich für Aufgaben rund um maschinelles Lernen und Künstliche Intelligenz (KI) einsetzen lässt. Ursprünglich entwickelte Google die Software für den internen Bedarf. Das Framework bietet vielfältige Einsatzmöglichkeiten und gestattet es, lernende neuronale Netze zu erstellen und diese zu verändern. Es zeichnet sich durch seine gute Skalierbarkeit aus und ist auf unterschiedlichen Systemen vom Smartphone bis zu Clustern mit vielen Servern betreibbar. Außerdem verweist Google auf verschiedenste trainierte Modelle welche für das Projekt verwendet werden. Tensorflow kann auf allen CPU's betrieben werden, oder mit Nvidia Grafikkarten, welche CUDA unterstützen.

- Installation in Python: pip install –upgrade gpu-tensorflow

Verwendet wird zusätzlich die API von Tensorflow (Models) für die Objekterkennung. In dieser wird eine Grundstruktur für das Trainieren eines CNN zur Verfügung gestellt, welche man je nach Verwendungszweck anpassen kann.

3.2.2 Benötigte Bibliotheken

Für das Arbeiten mit Tensorflow und CNN's benötigt man einige spezielle Python Bibliotheken. Diese werden im folgenden aufgeführt und erläutert. Die Installation geschieht hier über den P̈ackage Installerv̈on Python.

Pillow

Pillow ist eine Ïmaging Library f̈ür Python. Diese Bibliothek wird dafür genutzt um Bilddateien zu öffnen, verändern und speichern zu können. Pillow benötigen wir, um in unserem Künstlichen neuronalen Netz mit Bildern und Videos arbeiten zu können.

- Installation in Python: pip install –upgrade pillow

lxml

lxml ist eine Weiterführung der XML Sprache für Python. lxml erweitert dabei den Elementebaum von XML stark. Verwendet wird lxml dafür, um die Markierungen von Klassen und deren Positionen auszulesen, damit diese umgewandelt werden können. Durch die Arbeit mit Elementbäumen werden nur die benötigten Informationen herausgefiltert, wodurch wir eine schnellere Ausführung und bessere Arbeitszeiten erreichen.

- Installation in Python: pip install –upgrade lxml

Cython

Python ist im vergleich zu C eine recht langsame Programmiersprache, was die Ausführungzeit des Programmcodes wesentlich erhöht. Cython ist eine Programmbibliothek und auch Programmiersprache, welche die wesentlich langsame Programmiersprache Python erweitert und in der Ausführung beschleunigt. Dabei wird Python-Code in C kompiliert oder es kann externer code von C, eingebunden werden um ihn für die Ausführung zu nutzen. Das hat den Vorteil, das die Geschwindigkeit des Programms erhöht wird. Gerade für das Trainieren künstlicher neuronaler Netze, wird viel Rechenleistung benötigt. Das führt zu einer hohen Ausführungszeit des Programms. Um eine bessere Performance erreichen zu können greifen wir auf Cython zurück.

- Installation in Python: pip install –upgrade Cython

Jupyter

Jupyter oder auch J̈upyter Notebookïst eine open-source Web Applikation die es ermöglicht Dokumente welche Live Code, Abbildungen oder Text beinhalten zu teilen. Dabei kann man den Inhalt der Dateien im Browser aufrufen und ausführen. Verwendet wird diese Bibliothek, um die Lauffähigkeit des Neuronalen Netzes zu überprüfen.

- Installation in Python: pip install –upgrade jupyter

matplotlib

Die Programmbibliothek matplotlib ermöglicht es mit python, mathematische Darstellungen in der Python-Konsole oder im Code integriert anzufertigen und Visuell auszugeben. Matplotlib wird zum einen dafür verwendet die Erkennung, eines Objektes, auf einem Bild, an der jeweiligen Position anzuzeigen und zu beschreiben.

- Installation in Python: pip install –upgrade matplotlib

Pandas

Pandas ist eine Programmbibliothek von Python. Der Name stammt von P̈aneldata, was eine Bezeichnung von Datensätzen ist. Diese Bibliothek kann für die selektierung und Inditzierung von Daten verwendet werden. In dem künstlichen neuronalen Netz, wird pandas unter anderem für die Erhebung der XML Dateien der Trainingsdaten verwendet.

- Installation in Python: pip install –upgrade pandas

opencv-python

OpenCV ist ein Bildverarbeitungsprogramm, mit welchem einige Parameter von Bildern und Videos verändert werden können. OpenCV steht dabei für Öpen Computer Visionünd beschäftigt sich mit dem Maschinellen sehen. Dadurch, das dass zu entwickelnde CNN auf Basis von visuellen Daten arbeiten soll, welche mehr als ein Objekt gleichzeitig beinhalten kann, benötigen wir eine Aufteilung von Bildbereichen. Weitere Einsatzbereiche entstehen dann, wenn Beispielsweise die Helligkeit des Bildes nicht ausreicht. Hierbei kann die Helligkeit künstlich angepasst werden.

- Installation in Python: pip install –upgrade opencv-python

3.3 Trainingsdaten Generieren

Zu beginn ist ein un-trainiertes neuronales Netz noch "Dumm", da es bis zu diesem Zeitpunkt noch kein wissen darüber hat, welche Objekte es erkennen soll und welche Eigenschaften diese haben. Damit das zu entwickelnde Neuronale Netz weiß was es erkennen und klassifizieren soll, müssen zunächst Trainings und Testdaten generiert werden. Für das Training werden optimaler weise mehrere Tausend Trainings- und Testbilder benötigt, damit eine möglichst hohe Genauigkeit erreicht werden kann. Auch die Trainingsphase dauert in der Ausführung einige Stunden.

Wie in der Auswahl des Neuronalen Netz bereits beschrieben wird wegen der begrenzten Zeitressource ein trainiertes Netz von COCO verwendet. der Vorteil bei der Verwendung eines bereits trainierten Netzes besteht darin, das die Gewichtungen der frühen Layer gut eingestellt sind und diese nicht verändert werden müssen. Lediglich die obersten Layer sollen mit den eigenen Daten um-trainiert werden. Dadurch das die Gewichtungen der untersten Layer trainiert sind werden weniger Trainings- und Testdaten benötigt, um eine hohe Genauigkeit des Neuronalen Netzes zu erreichen.

3.3.1 Funktionsweise

Wie gerade schon angeschnitten, bekommen bei Netzen welche neu trainiert werden jedes Neuron zunächst ein zufälliges Anfangs-Gewicht und Schwellwert. Beim starten der Trainingsphase werden Daten in das Netz eingeführt. Jedes Neuron verarbeiten nun alle Eingangssignale mit der zugewiesenen Gewichtung und geben die Ergebnisse an das nächste Neuron weiter.

Im Output-Layer wird nun das Ergebnis der Berechnungen ermittelt. Das Ergebnis des Netzes wird mit der Tatsächlichen Bezeichnung des Objektes verglichen. Lag das Netz bei

der Bestimmung falsch wird zunächst der entstandene Fehler Berechnet und die Neuronen werden dementsprechend angepasst. Dabei werden die Gewichtungen der Neuronen so stark verändert je mehr Einfluss diese auf das resultierende Ergebnis hatten. Dieser Vorgang wird nun viele male wiederholt. Das ist die kurze Fassung, wie ein Künstliches Neuronales Netz lernt.

3.3.2 Auswahl der Klassen

Bei der Auswahl der Klassen wurde darauf geachtet, die Genauigkeit des Netzes zu erkennen, indem Klassen ausgewählt werden, welch eine gewisse Ähnlichkeit aufweisen jedoch unterschiedliche Bezeichnungen haben. Beispiele sind hier Flaschen, Tetra Packs oder auch Schachteln.

- ID: 1 Name: 'Bier, Flasche'
- ID: 2 Name: 'Wasser, Flasche'
- ID: 3 Name: 'Milch, Packung'
- ID: 4 Name: 'Orangensaft, Packung'
- ID: 5 Name: 'Brunch, Aufstrich'
- ID: 6 Name: 'Margarine'

3.4 Trainieren

Damit das zu entwickelnde Netz für die ausgewählten Klassen verwendet werden kann, muss ein großer Datensatz generiert werden. Wie im obrigen Abschnitt beschrieben, werden Klassen ausgewählt, welche eine gewisse Ähnlichkeit aufweisen. Dadurch soll überprüft werden, ob das neuronale Netz auch Objekte mit nur kleinen Unterschieden auseinanderhalten kann.
Die ausgewählten Objekte für die Klassifizierung, sind jeweils zwei Packungen, Flaschen und Aufstriche. Wie die Einteilung der Klassen im Detail aussehen wird in der Folgenden Tabelle gezeigt.

Tabelle 2: Ausgewählte Klassen und Struktur

Klasse	Bezeichnung	Bereich
1	Milch - Packung	1-99
2	Orangensaft - Packung	100-199
3	Wasser - Flasche	200-299
4	Bier - Flasche	300-399
5	Brunch - Aufstrich	400-499
6	Margarine	500-599
7	Gemischt	600-699

Zunächst müssen Testdaten in Form von Bildern generiert werden, dafür wurden zum einen, Bilder von Online-Suchmaschinen, je nach Verwendbarkeit für das Netz heruntergeladen. Dabei sollte das Bild eine Größe von 200 KB nicht übersteigen, damit das Training erfolgreich verläuft und nicht zusätzlich mehr Zeit benötigt. Des weiteren wurden Objekte der aufgeführten Klassen, in verschiedensten Winkeln und Lichtverhältnissen Fotografiert. Die aufgenommenen Bilder übersteigen mit 3-5 MB die maximale Größe der Testdaten. Aus diesem Grund, werden die Bilder mit einem Programm um den Faktor 0.2 verkleinert. Verwendet wird für die Komprimierung ein in Python geschriebenes Programm (reziser.py).

Damit die API von Tensorflow weiß, welche Klassen auf dem Bild zu sehen sind und

wo sich diese im Bild befinden, müssen die Bilder gelabelt werden. Das Programm labelImg.py wird dafür verwendet. Hier wird jedes Bild mit den Positionsdaten und und der Bezeichnung markiert und in einer XML Datei abgespeichert. Nach dem Markieren werden Die Daten in Trainings und Testdaten aufgeteilt. 80 Prozent kommen in den Trainingsordner und 20 Prozent in den Testordner. Für die Durchführung des Trainings werden die Informationen der einzelnen XML Dateien, in zwei CSV Dateien zusammengefasst. Diese Dateien beinhalten Tabellen in denen alle Notwendigen Informationen zu den Bildern aufgeführt werden. Verwendet wird dafür das Programm xml_to_csv.py.

Nachdem alle Bilddaten gelabelt wurden und die CSV Dateien für die Trainings- und Testdaten erstellt wurden, kann das Training beginnen. Für das Training benötigen wir die Trainingsdaten, welche im vorhinein erstellt wurden, eine Labelmap (labelmap.pbtxt) in welcher unsere Ausgewählten Klassen eingetragen sind. Außerdem brauchen wir die konfigurations Datei des verwendeten Netzes (faster_rcnn_inception_v2_coco). Das Training wird mit Ausführung des training.py Programms gestartet. Ab diesem Zeitpunkt, benötigt das Training viel Zeit.

Zusätzlich Parameter:
–logtostderr
–train_dir=training/
Dieser Parameter gibt den Pfad zum Ordner an in welchem die Trainingsergebnisse Gespeichert werden sollen.
–pipeline_config_path=training/faster_rcnn_inception_v2_coco.config
Dieser Parameter gibt den Pfad zu der benötigten config Datei an

Damit die Ergebnisse so gut wie möglich ausfallen, wird das Netz ca. 10 Stunden, mit der in Kapitel? beschriebenen Konfiguration Trainiert. Dabei schafft das Programm 200.000 Trainings- und Testdurchläufe und ist kurz vor Abschluss der Trainingsphase bei einem durchschnittlichen 'loss' von 0,03. Alle 5 Minuten werden die Zwischenstände gespeichert. Aus dem letzten Standpunkt des Trainings wird mit hilfe des export_inference_graph.py ein Sogenannter 'frozen inference graph' generiert, welcher unser Künstliches neuronales Netz darstellt.

Mitgegebenen Parameter:
–input_type image_tensor
–pipeline_config_path legacy/training/faster_rcnn_inception_v2_coco.config
Gibt den Pfad zur config Datei an.
–trained_checkpoint_prefix legacy/training/model.ckpt-XXXX

Gibt den Pfad zur letzten Speicherstand des Trainings an.
–output_directory inference_graph
Bestimmt den Ort an welchen der Graph exportiert wird.

Ab diesem Zwischenstand, kann das Neuronale Netz für die Objekterkennung verwendet werden. Im weiteren Unterpunkt, wird das Netz mit Ungelabelten Bildern getestet. Dabei soll herausgestellt werden, wie zuverlässig das Netz ist und ob es alle Objekte erkennt.

4 Testergebnisse

Wie im vorherigen Kapitel schon beschrieben wurde, soll in diesem Kapitel die Zuverlässigkeit sowie Geschwindigkeit, des neuronalen Netzes geprüft und ausgewertet werden. Dabei wurden Testbilder verwendet, welche nicht im Trainings- und Testdatensatz vorhanden waren. Es soll sichergestellt werden, dass das Netz die Objekte nicht durch das Training kennt.
Tensorflow bietet mit dem Tool Tensorboard die Möglichkeit den fortschritt des Trainings grafisch darzustellen. In der folgenden Grafik wird der auftretende 'Loss' in Verhältnis zu der Trainingszeit gestellt.

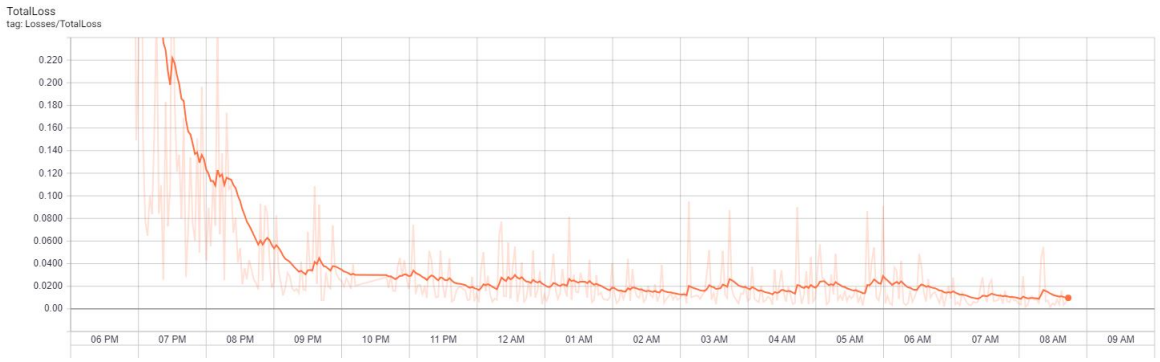

Abbildung 5: Loss Graph des umtrainierten inceptionV2

Der 'Loss' stellt dabei keine Prozentangabe dar, sondern ist eine Zusammenfassung der Fehler, die für jedes Beispiel in Trainings- oder Validierungssätzen gemacht wurden. Je kleiner der 'Loss dabei ist, desto höher ist die Wahrscheinlichkeit. Um die Grafik zu veranschaulichen, wurden kurzfristige Schwankungen geglättet. Zu beginn wird festgestellt, das die ersten 4 Stunden des Trainings den größten fortschritt aufweisen. In dieser Zeit, ist der 'Loss' von anfangs 2.0 auf 0.05 gesunken. Dabei wurden rund 60.000 Trainingseinheiten durchlaufen.
Die ersten Testergebnisse sehen gut aus. Alle beinhaltenden Objekte konnten erkannt und zugewiesen werden. Diese wurden auch mit einer hohen Wahrscheinlichkeit identifiziert. Lediglich bei der 'Orangensaft - Packung' hat das Netz einen Fehler gemacht. Diese wurde als Milch erkannt und das auch mit einer recht hohen Wahrscheinlichkeit. Zurückzuführen wäre das durch die hohe Ähnlichkeit der Formen.

Aus diesem Grund, soll überprüft werden, ob beim aufsetzen des Neuronalen Netzes Fehler gemacht wurden. Außerdem soll ein weiteres Vor-trainiertes Netze mit dem Datensatz getestet werden, Genauigkeit untereinander zu vergleichen. Das dafür ausgewählte Netz ist das faster_rcnn_resnet101_coco. Dieses Modelle wurden aus der Tabelle? ausgewählt, da die mittlere durchschnittliche Genauigkeit.

Abbildung 6: Erster Test des Neuronalen Netzes

Das zweite Modell welches trainiert wurde ist das faster_rcnn_resnet101_coco mit einer mAP von 32 dafür ist es aber auch wesentlich langsamer. Für einen Trainingsdurchlauf benötigt das Modell mit rund 600ms im durchschnitt, mehr als doppelt so lange wie das faster_rcnn_inception_v2_coco. Es benötigt dafür aber auch nur die Hälfte an Durchläufen, um eine ähnlich niedrigen 'Loss' zu erreichen.

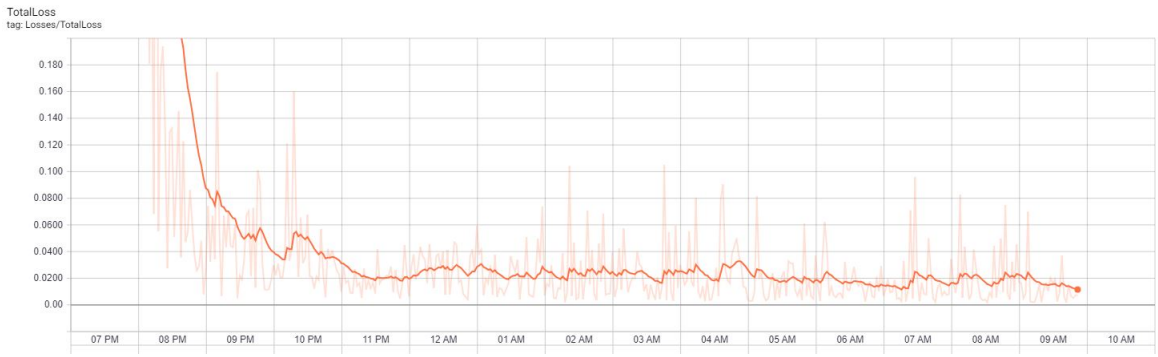

Abbildung 7: Loss Graph des umtrainierten resnet101

Testergebnisse der einzelnen Klassen zeigen auch das die Genauigkeit besser ist. Der Orangensaft welcher nicht identifiziert wurde, konnte jetzt erreicht werden. Dabei gab es wiederum das Problem das dieser je nach Winkel nicht immer richtig eingeordnet wurde. Um die Stabilität der ausgaben weiter zu erhöhen, sollen die Trainingsdaten weiter angehoben werden und das letzte Modell welches die höchste mAP hat soll auf den Datensatz trainiert werden.

Weitere Test mit dem Trainierten Netz haben ergeben, das die Erkennung der Orangensaft Packung funktioniert, solange das Bild und die Oberflächen der Objekte nicht zu dunkel werden. Das Bild (Bild) zeigt alle Trainierten Objekte mit welchen das Modell

faster_rcnn_resnet101_coco trainiert wurde. Hier kann man erkennen das alle Klassen richtig zugeordnet werden konnten.

Abbildung 8: Objekt erkennung des optimierten resnet101

Abgesehen von den schwächen in der Erkennung ähnlicher Objekte hat das generieren eines Neuronalen Netzes Funktioniert. Im nächsten Kapitel sollen die Ergebnisse des Projektes, mit den Anforderungen verglichen und ausgewertet werden.

5 Fazit

Wie im vorherigen Kapitel beschrieben, hat die Entwicklung eines Künstlichen Neuronalen Netzes gut funktioniert. Lediglich bei ziemlich ähnlichen Objekten, gab es Probleme mit der richtigen Identifizierung. Um ein finales Fazit geben zu können soll in diesem Kapitel auf die Anforderungen eingegangen werden, welche zu Beginn des Projektes festgelegt wurden.

5.1 Bewertung durch Anforderungen

Die ersten Zwei funktionalen Anforderungen konnten erfüllt werden. Lediglich die Genauigkeit konnte zunächst nicht zu voller Zufriedenheit erfüllt werden. Die meisten Objekte können bei schlechteren Bild- und Lichtverhältnissen erkannt werden, lediglich bei zu ähnlichen Objekten kann die Erkennung fehlerhaft sein, wie man bei der Milch und dem Orangensaft beobachten kann. Mit mehreren Trainingsdaten und einer Licht Korrektur könnte dies allerdings verbessert werden.

Die organisationalen Anforderungen konnten auch zum Großteil erfüllt werden. Klassen können unkompliziert erweitert werden, indem Testdaten der gewünschten Klasse Generiert und gelabelt werden. Das trainieren und ausführen des Neuronalen Netzes ist zwar im Moment nicht mit einer Datei ausführbar, aber die notwendigen schritte sind nicht aufwendig. Die Installation aller Notwendigen Module und Bibliotheken kann durch die Eingabeaufforderung oder der shell ausgeführt werden. Lediglich für Anaconda muss man eine manuelle Installation durchführen. Das Netz kann auf jedem Endgerät mit der benötigten Entwicklungsumgebung betrieben werden. Für mobile Endgeräte ist das Netz jedoch wegen der benötigten Rechenleistung nicht geeignet.

Bei den qualitativen Anforderungen kann man nicht jede erfüllen, weil je höher die Genauigkeit des Netzes ist, desto länger benötigt es um ein Bild auszuwerten. Umgekehrt sinkt die Genauigkeit umso schnelle das Netz arbeitet. Es wurde also darauf geachtet das es bei beiden Punkten solide Ergebnisse erzielt. Der Schwerpunkt lag aber auf der Genauigkeit des Netzes. Leicht verdeckte Objekte können auch bis zu einem Gewissen Punkt erkannt und richtig zugeordnet werden. Die Genauigkeit kann beispielsweise, durch mehrere Trainingsdaten erhöht werden.

5.2 Bewertung der Netze

Das Neuronale Netz funktioniert größtenteils wie gewollt, nur bei zwei Klassen kann nicht immer Korrekt unterschieden werden. Das liegt hauptsächlich an der hohen Ähnlichkeit der zwei Objekte. Die Genauigkeit könnte durch eine höhere Anzahl an Trainings und Testdaten verbessert werden. Außerdem könnte eine Angleichung der Lichtverhältnisse per Bildverarbeitung dafür sorgen, dass die Genauigkeit des Netzes steigt. Zwischen den getesteten Modellen wurde festgestellt, das die Übereinstimmung der Objekte einen wesentlichen unterschied in der Genauigkeit aufweisen, sodass eine etwas höhere Antwortzeit für den Verwendungszweck kein großes Problem darstellt.

Abschließend, kann das Projekt als Erfolgreich angesehen werden, mit kleinen Optimierungen.

6 Aussicht

Die Entwicklung eines neuronalen Netzwerkes hat gezeigt, das es Möglich ist den Inhalt Regal-Systeme mit Hilfe von Objekterkennung zu Identifizieren. Dabei ist aber auch aufgefallen, das mehr Trainingdaten benötigt werden je Ähnlicher sich die Klassen sind. Das bedeutet das für das geplante System eine große Anzahl an Trainingsdaten generiert werden müssen. Grundsätzlich ist es aber möglich mit den Ergebnissen weiter Arbeiten zu können.

Für das Geplante System wären bis auf die Optimierung des Künstlichen neuronalen Netzes, die Systemschnittstellen welche zum einen das Bild vom Regal aufnimmt, verarbeitet und an das neuronale Netz weiter leitet und zum anderen müssen die Informationen der Bilderkennung an eine Datenbank übergeben werden, welche den Inhalt des Regals speichert. Darüber hinaus müsste eine Anwendung entwickelt werden, mit welcher der Benutzer den Inhalt des Regals managen kann.

Die Optimierung soll aber Für die Nächsten Schritte im Vordergrund stehen. Für die anstehende Bachelorarbeit, soll die Verbesserung des Neuronalen Netzes im Vordergrund stehen. dabei gibt es drei wichtige Punkte welche behandelt werden müssen. Zum einen soll die Genauigkeit der Identifizierung bei ähnlichen Objekten erhöht werden. Außerdem sollen Bildaufnahmen so verarbeitet werden, das Lichtverhältnisse gleich sind oder eine Veränderung dieser Keinen Einfluss auf die Ergebnisse haben. Zusätzlich soll die Auswirkung schlechter Bildverhältnisse ermittelt und Verbessert werden.