Digital Business University of Applied Sciences
Studiengang: Data Science and Business Analytics (B. Sc.)
Modulbezeichnung: SPII-4, Forschungsprojekt Artificial Intelligence & Machine Learning
Dozent: Ashok Kaul

Studienarbeit

Apfel oder Birne?
Ein Vergleich Neuronaler Netzwerke zur Klassifizierung von Früchten

Eingereicht von: Conny Brintzinger
Matrikelnummer: 190044
Datum: 05.08.2023

Inhaltsverzeichnis

Abbildungsverzeichnis

1. Einführung

Ob Wocheneinkauf oder schnell noch ein paar Snacks und Getränke für den Abend, der Gang in der Supermarkt gehört zum Alltag und nimmt für Verbraucher einen hohen Stellenwert als zentrale Anlaufstelle zur Versorgung mit Lebensmitteln ein. Auf rund 16 Millionen Quadratmetern Verkaufsfläche erwirtschafteten deutsche Supermärkte in 2021 rund 79,1 Milliarden Euro Jahresumsatz. In den letzten 13 Jahren hat insbesondere der Warenbereich „Frische“ an Bedeutung gewonnen und beinhaltet heute doppelt so viele Artikel. (Apúntateuna, 2023).

Um den Kunden ein noch besseres Einkaufserlebnis zu bieten, lange Wartezeiten an der Kasse zu ersparen und weitere nützliche Services anzubieten, kommt immer öfter Künstliche Intelligenz (KI) zum Einsatz. Es ist zu erwarten dass kassenlose 24/7-Stores, die ganz ohne Personal auskommen, zukünftig zum Alltag gehören. Bereits heute gibt es in Deutschland 80 dieser sogenannten „Smart Stores“, in denen Kameras erfassen, welche Produkte in den Korb gelegt werden und diese via App bei Verlassen des Stores automatisch abrechnen (Dittrich, 2023).

Während abgepackte Produkte durch Barcodes gekennzeichnet sind, müssen unverpacktes Obst und Gemüse bisher durch einen Mensch identifiziert und zugeordnet werden. An der Kasse passiert das häufig durch Zahlencodes, die das Kassenpersonal auswendig lernen oder nachschlagen muss. Bei Selfservice-Kassen wird der Kunde hingegen durch ein bebildertes Menü geführt und wählt dort das entsprechende Obst oder Gemüse aus. Diese Praxis ist zeitraubend und fehleranfällig. Zudem entsteht durch solche Bedienfehler ein fehlerhaftes Inventar.

Um Objekte, wie Obst oder Gemüse, erkennen und klassifizieren zu können, stehen unterschiedliche Machine Learning Frameworks zur Verfügung. Convolutional Neural Networks (CNN) bieten im Vergleich zu herkömmlichen Neuronalen Netzwerken eine höhere Effizienz bei geringerem Speicher- und Rechenbedarf sowie eine hohe Robustheit und können daher als State-of-the-Art in der Bild- und Audioverarbeitung angesehen werden (Oldenthal, 2019).

Im Folgenden werden zwei unterschiedliche CNN etabliert und deren Ergebnisse bei der Klassifizierung von Obst gegenübergestellt.

2. Initialisierung & Daten

Die Datenbasis für das Training der Modelle liefert der „Fruits 360“ Datensatz (Oltean, 2017). Er enthält insgesamt 90.483 Fotos von 131 Klassen von Obst- und Gemüsesorten. Entstanden sind die Fotos, indem das Obst / Gemüse auf einer sich langsam drehenden Welle vor weißem Hintergrund positioniert und mit einer Logitech C920 Kamera jeweils kurze Filme aufgenommen wurden. Auf diese Art wurde eine Vielzahl von Einzelbildern generiert, die die Objekte von unterschiedlichen Seiten zeigen. Aufgrund der unterschiedlichen Beleuchtungsbedingungen und dem resultierenden Schattenwurf wurden die Fotos anschließend freigestellt. Zuletzt wurden die Bilder zentriert und auf eine Größe von 100x100 Pixel komprimiert. Die Herkunft der Daten wird als vertrauenswürdig eingeschätzt. Nach einer ersten visuellen Prüfung kann die Qualität der Bilder als gut bezeichnet werden.

Um den Trainingsumfang zu begrenzen, wurden 5 Klassen von Obst (5284 Bilder) ausgewählt, die klassifiziert werden sollen:

- Apple (2134 Bilder)
- Apricot (492 Bilder)
- Limes (490 Bilder)
- Orange (479 Bilder)
- Pear (1689 Bilder)

Da für Birnen und Äpfel unterschiedliche Sorten mit charakteristischen Merkmalen in Farbe, Form und Muster differenziert werden, liegen für diese beiden Klassen deutlich mehr Trainingsdaten vor als für Limetten, Aprikosen und Orangen. Zum Training der Modelle werden 90% des Datensets benutzt, während 10% der Bilder zur Validierung dienen.

In Anlehnung an den „Fruits“ Datensatz wurde ein zweites, selbst fotografiertes Test-Datenset entwickelt, um die Modelle mit Fotos anderer Früchte der gleichen Klassen vor eine realistische Herausforderung zu stellen. Die Früchte wurden mit einem Google Pixel 6 Pro vor weißem Papier-Hintergrund fotografiert, zentriert und im Format 100x100 Pixel abgespeichert. Im Unterschied zum Trainingsdatenset wurden die Fotos nicht freigestellt. Daher haben die Bilder einen Hintergrund, der in Grautönen verläuft und den Schattenwurf der Früchte zeigt. Wie beim Trainingsdatenset erfolgte die Markierung der zugehörigen Labels, indem die Fotos in entsprechend benannten Ordnern abgelegt wurden.

3. Modellierung

In den letzten zehn Jahren haben sich Neuronale Netzwerke (NN) insbesondere in Bereichen der digitalen Sprach- und Bilderkennung gegenüber früheren Machine Learning Verfahren durchgesetzt. Durch die inkrementelle, schichtweise Vorgehensweise, bei der zunehmend komplexere, abstraktere Repräsentationen entwickelt werden, und das gleichzeitige Erlernen der Merkmale entlang der Layer können Aufgaben von hoher Komplexität einfach, skalierbar und vielseitig und gelöst werden (Chollet, 2018).

CNN ermöglichen durch ihre integrierten Dense-Layer ein translationsinvariantes Erlernen von Mustern. Das heißt, ein einmal erlerntes Muster kann überall im Bild wiedergefunden werden. Darüber hinaus können räumliche Hierarchien erkannt und zunehmend komplexere und abstraktere visuelle Konzepte erlernt werden (Chollet, 2018). Zur Klassifizierung des „Fruits“ Datensatzes wurden zwei unterschiedliche CNN-Modelle trainiert und getestet:

Model 1 (Notebook: EDA_and_CNN_Fruits-Dataset.ipynb) entspricht einem gängigen, auf Forschung mit ähnlichen Daten basierenden Framework. Es beinhaltet drei Convolutional-Schichten (32, 64 und 128 Filter) mit dazwischen liegenden Pooling-Schichten (MaxPooling), einem Flatten-Layer und abschließenden Dense-Layern. Als Aktivierungsfunktion wurde in den Convolutional-Schichten wie auch im letzten Dense-Layer Relu verwendet. Um die bestmöglichen Gewichte für das CNN zu finden, wurde Adam als Optimierungsfunktion gewählt. Die Verlust-Funktion, also der Fehler zwischen Vorhersage und vorgegebenem Zielwert, wird durch Categorical Crossentropy überwacht.

In **Modell 2** (Transfer Learning_Inference_Fruits-Dataset.ipynb) wurde ein vortrainiertes CNN genutzt, um von den in der Faltungsbasis erlernten Merkmalen zu profitieren und diese auf die „Fruits“ Daten zu übertragen. Als vortrainiertes Model wurde efficientnet_v2_b0 (Tensor Flow Hub, 2022) ausgewählt, da es im Vergleich zu State-of-the-Art-Modellen sehr leichtgewichtig ist und schnelles Training ermöglicht (V. Le & Tan, 2021). Das Training erfolgte auf der ImageNet Datenbank und somit auf ca. 14 Millionen farbigen Bildern im RGB-Format. Wie im anderen Modellen wurden auch hier Adam als Optimierungsfunktion und Categorical Crossentropy als Verlustfunktion ausgewählt.

4. Ergebnisse

Zur Evaluation der Ergebnisse wurde zunächst die Accuracy herangezogen, um den Anteil der korrekten Vorhersagen an der Gesamtheit der Vorhersagen zu ermitteln:

Model 1 Accuracy: 1,00

Model 2 Accuracy: 1,00

Sowohl **Model 1** als auch **Model 2** erreichen bereits in der dritten Epoche eine perfekte Accuracy von 100%. Während es Ziel der Image-Klassifizierung ist, eine möglichst hohe Trefferquote zu erreichen, sprechen diese Werte eher für eine Überanpassung des Models an die Trainingsdaten. Wie beschrieben, sind die Trainingsdaten aus den Einzelbildern eines Films der langsam bewegten Frucht entstanden. Daher weichen die Ansichten nur minimal voneinander ab. Jedes der Bilder im Test-Datensatz hat mehrere Bilder in den Trainingsdaten, die fast identisch sind. Daraus erklärt sich die hohe Trefferquote.

Die folgende *Abbildung 1* zeigt einen Inference-Plot von **Model 2**. Alle Früchte wurden der richtigen Klasse zugeordnet.

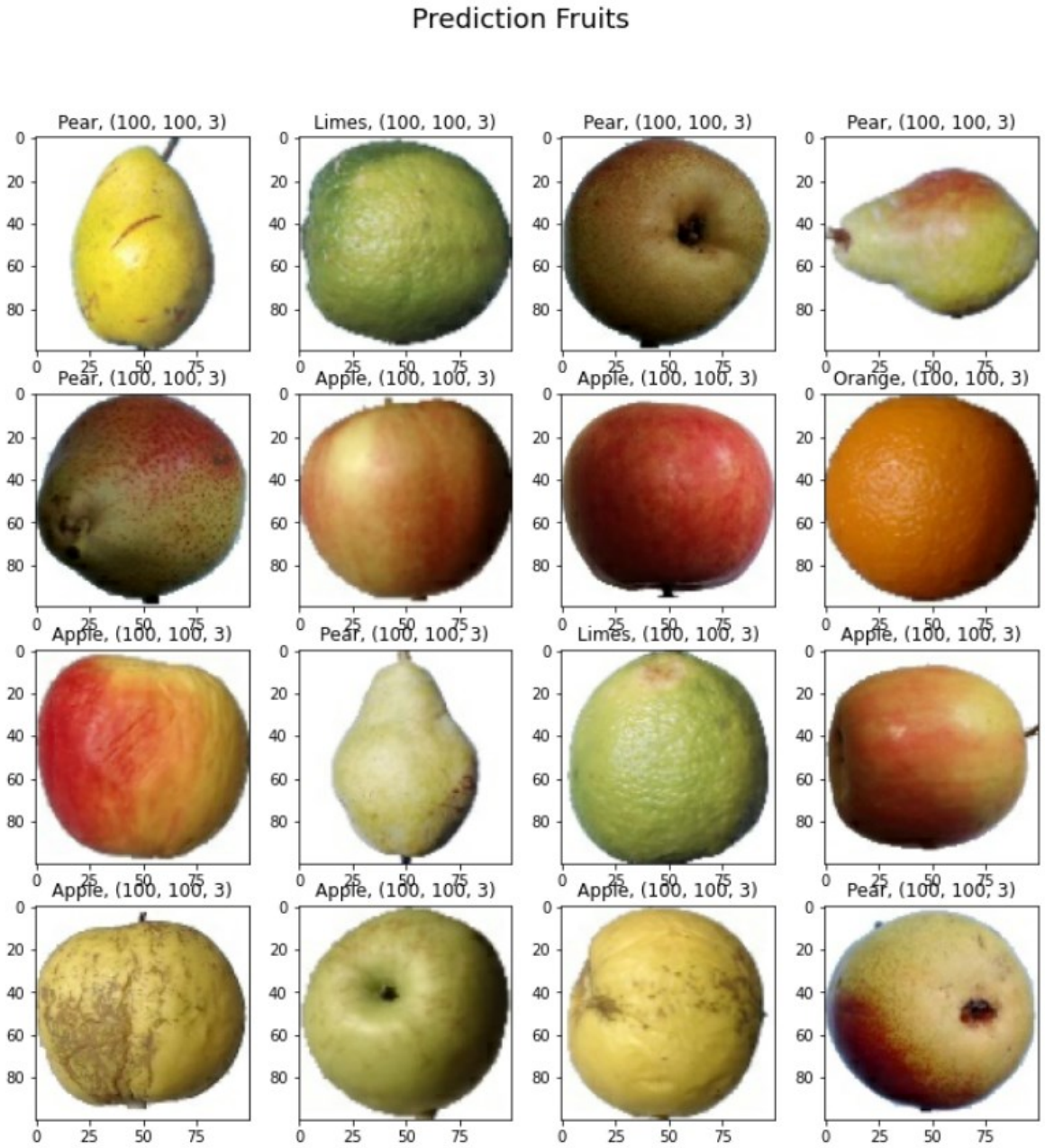

Abbildung 1: Inference Plot Transfer-Learning-Model

Um zu testen, inwiefern die beiden Modelle tatsächlich die Eigenschaften der verschiedenen Früchte unterscheiden und richtig zuordnen können, wurden die Modelle mit neu erstellten Fotos der fünf Früchte- Klassen (*Apple, Apricot, Limes, Orange, Pear*) konfrontiert.

Dieser Test führte zu stark abweichenden Ergebnisse und brachte folgende Ergebnisse:

Model 1 Accuracy: 0,25

Model 2 Accuracy: 0,25

Mit 25% Accuracy liegen beide Modelle in ihrer Treffsicherheit nur geringfügig über einer zufälligen Verteilung von 20%, die sich aus den 5 Klassen ergibt. Die folgende *Abbildung 2* zeigt einen Inferenz-Plot von **Modell 2** mit den neu erstellten Fotos.

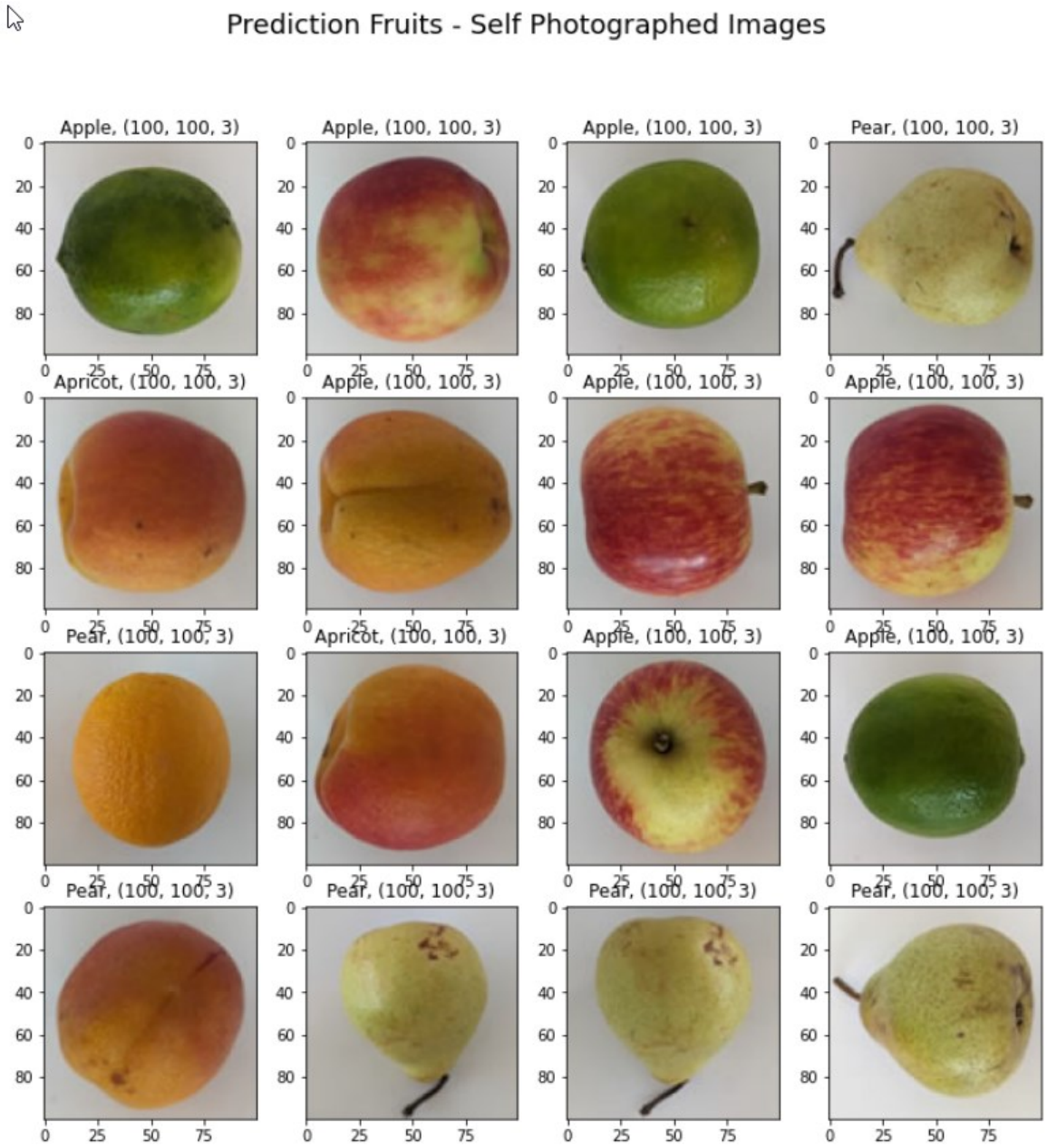

Abbildung 2: Inference Plot Transfer-Learning-Model mit neu erstellten Fotos

Man kann sehen, dass Äpfel häufig richtig klassifiziert wurden. Auch Birnen wurden gut erkannt, während Limetten als Äpfel kategorisiert wurden.

Die Inference Ergebnisse entlang der einzelnen Klassen können mittels einer Confusion Matrix anschaulich dargestellt werden. Wie in *Abbildung 3* zu erkennen, gibt es beim Transfer Learning **Model 2** die meisten abweichenden Labels in den Klassen *Apple* und *Pear*, wobei es sich häufig um Verwechslungen der beiden Früchte handelt.

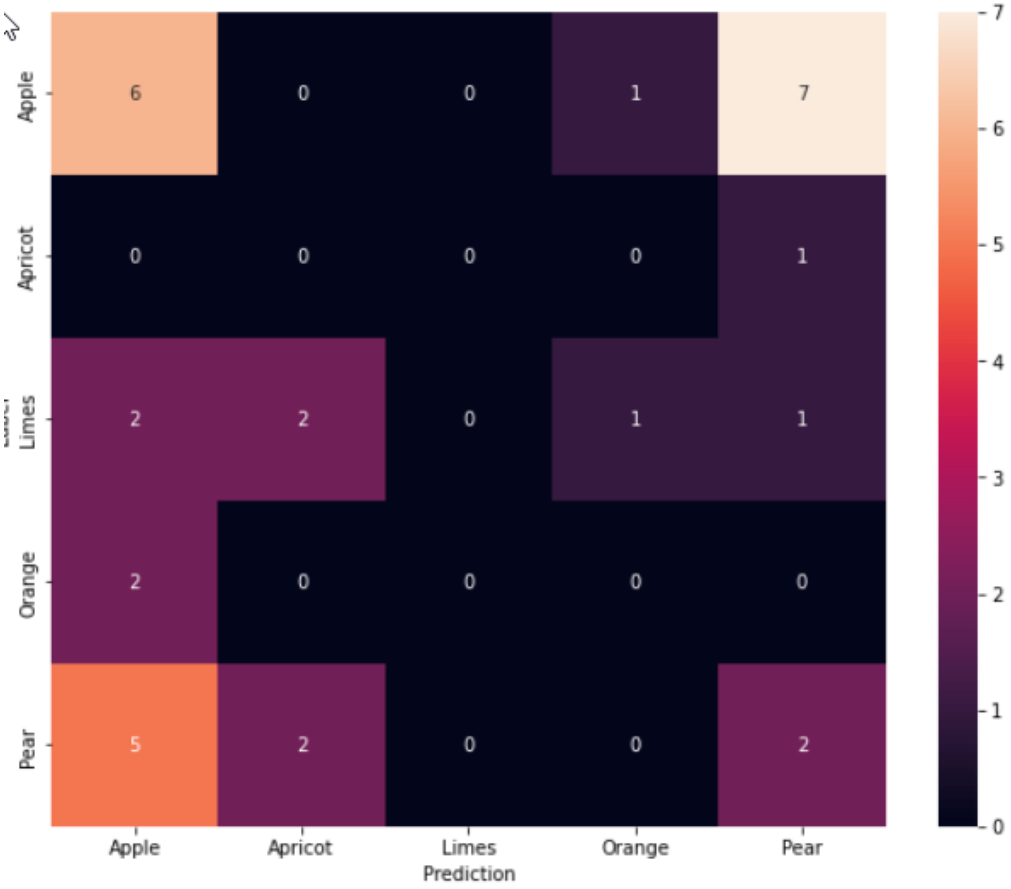

Abbildung 3: Confusion Matrix Transfer-Learning-Model mit neu erstellten Fotos

Die Klasse *Apricot* wurde 4mal vergeben, es handelte sich jedoch 2mal um *Limes* und zweimal um *Pear. Orange* wurde zweimal klassifiziert, ebenfalls in beiden Fällen falsch und *Limes* wurde als Label gar nicht vergeben.

Für das CNN **Model 1** ergeben sich ähnliche Schwächen, wie die Confusionmatrix in *Abbildung 4* verdeutlicht.

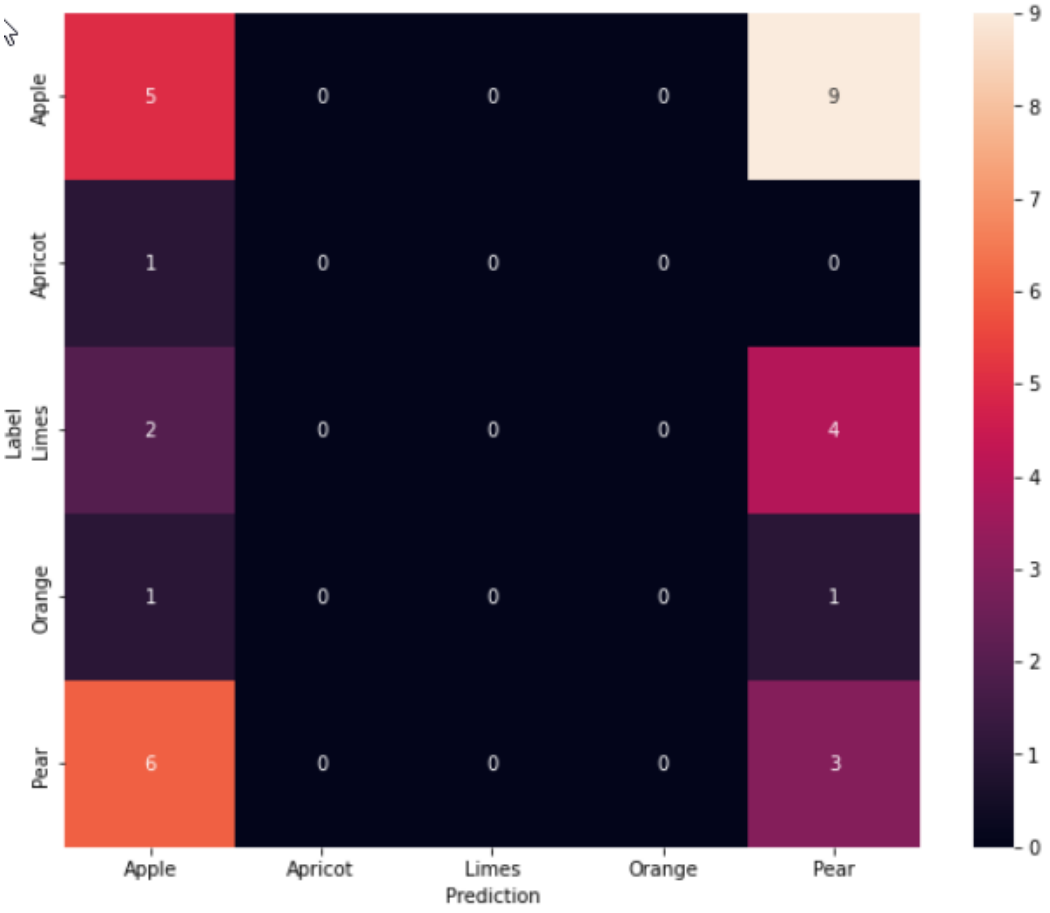

Abbildung 4: Confusion Matrix CNN-Model mit neu erstellten Fotos

Im Unterschied zu **Modell 2** wurden die Klassen *Apricot*, *Limes* und *Orange* nie vergeben, sondern lediglich zwischen *Apple* und *Pear* variiert. Dabei wurde *Apple* 5mal richtig und 9mal falsch klassifiziert. *Pear* hingegen wurde 3mal richtig erkannt und 6mal fälschlicherweise als *Apple* klassifiziert.

Obwohl Transfer Learning-Modelle bekannt sind für ihre hohe Performance, konnte das **Model 2** die Werte von **Model 1** nicht übertreffen. Vortrainierte Modelle sind in der Lage, selbst mit wenigen Daten gute Ergebnisse zu erzielen. Der Testdatensatz sollte mit 4279 Bildern eine geeignete Basis liefern.

Für das unverhältnismäßig gute Abschneiden beider Modelle mit den Testdaten aus dem „Fruits“ Datenset als auch die schlechte Performance mit den selbst erstellten Fotos können folgende Gründe angeführt werden:

1. Minimal variierende Fotos der gleichen Frucht im Trainingsdatenset

Da die Bilder der einzelnen Früchte im „Fruits“ Datensatz nur minimal variieren, haben die Modelle vermutlich stärker auf das Lernen von Unterschieden zwischen den Bildern fokussiert als auf Merkmale, die die Früchte voneinander unterscheiden. Daher konnte für den „Fruits“ Datensatz eine unrealistisch hohe Accuracy von 100% erreicht werden, während die Performance bei den selbst fotografierten Fotos mit 25% Accuracy deutlich hinter den Erwartungen zurückblieb.

2. Abweichende Kamera, Hintergrund und Früchte

Die Modelle wurden mit Fotos trainiert, auf denen die gleichen Klassen von Früchten abgebildet waren, aber nicht dieselbe Frucht wie bei den selbst fotografierten Bildern, wodurch sich Unterschiede im Aussehen ergeben. Zudem entstehen Abweichungen durch Verwendung einer anderen Kamera und Komprimierung sowie durch die Tatsache, dass die eigenen Fotos nicht freigestellt wurden, wie im Test-Datensatz. Insofern unterscheiden sich die selbst fotografierten Bilder in mehreren Aspekten stark von den Trainingsdaten, was zu fehlerhaften Klassifizierungen führt.

3. Dataset Imbalance

Das Trainings-Datenset war nicht ausgewogen. Insbesondere die Klassen *Apricot*, *Limes* und *Orange* waren mit je 492 / 490 / 479 Bildern stark unterrepräsentiert. Die meisten Fotos entsprachen mit 2134 Bildern der Klasse *Apple*. Auch *Pear* war mit 1689 Bildern häufig vertreten. Diese Unausgewogenheit spiegelt sich in den Inferenz-Ergebnissen wider. Beide Modelle haben deutlich häufiger die Klassen Äpfel oder Birnen vergeben. Diese Verzerrung ist als direkte Folge der unausgewogenen Klassenverteilung im Trainingsdatenset zu sehen.

5. Schlussfolgerungen

Beide Modelle zeigen mit einer Accuracy von 100% außergewöhnlich gute Ergebnisse mit Testdaten, die dem „Fruits“ Datenset vor Training der Modelle zur Evaluation entnommen wurden. Das erklärt sich aus der starken Ähnlichkeit der Bilder, die lediglich in kleinen Details variieren.

Beim Test der Modelle mit eigenen Fotos der Früchte zeigen beide Modelle mit jeweils 25% Accuracy eine unzureichende Performance. Um die Accuracy zu erhöhen, stehen grundsätzlich zwei Stellschrauben zur Verfügung: die Trainingsdaten oder das Modell selbst. Letzteres kann durch Hyperparameter-Tuning, wie Random Search verbessert werden. Ein Experimentieren mit den Parametern, wie Anzahl der Schichten und Neuronen, Änderung der Lernrate und Epochen sowie der Aktivierungsfunktion und des Optimizers kann die Performance weiter verbessern. Auch Regularization-Ansätze, wie Dropout oder Early Stopping können zu besseren Klassifizierungs-Ergebnissen beitragen.

Weit vielversprechender erscheint in diesem Fall ein Ansatz zur Steigerung der Trainingsdaten-Qualität. Durch mehr qualitativ hochwertige, gut ausbalancierte Trainingsdaten können die Modelle deutlich verbessert werden. Die Trainingsdaten sollten sich dabei so nah wie möglich am späteren Einsatz des Modells orientieren, also eine hinreichende Menge an realen Daten mit unterschiedlichen Hintergründen beinhalten. Ein Ausbalancieren der Klassen innerhalb des Trainingsdatensets kann die gesehene Verzerrung beseitigen und zu besseren Ergebnissen führen.

Anhand der getesteten Modelle konnte demonstriert werden, welches Potenzial CNN-Frameworks im Bereich Computer Vision und insbesondere bei der Klassifizierung von Bilddaten entfesseln, sofern beim Training auf hohe Datenqualität sowie geeignete Modelle und Parameter Wert gelegt wird. Die Klassifizierung von Früchten bietet eine Vielzahl an Ansatzpunkten, um die Customer-Experience in Supermärkten zu verbessern. Sowohl an herkömmlichen Kassen als auch an den immer häufiger anzutreffenden Selfservice-Kassen, an denen Kunden ihre Waren selbst Scannen und Bezahlen, können Effizienz und Komfort deutlich erhöht werden, wenn die auf der Waage platzierten Früchte per Kamera und verknüpfter KI identifiziert werden anstatt aufwendig im Menü zu suchen oder entsprechende Zahlencodes händisch einzutippen. Darüber hinaus können zusätzliche Kundenservices angeboten werden, wie nützliche Informationen zu den jeweiligen Produkten, Rezeptvorschläge oder der kürzeste Weg zu passenden Produkten. Besonders große Bedeutung kommt der Objekterkennung bei kassenlosen Märkten zu, welche aktuell von vielen großen Lebensmittel-Ketten getestet werden. Im Markt verteilte Kameras erfassen, welche Produkte der Kunde in den Einkaufwagen legt und rechnet diese automatisch über eine App ab, wenn der Kunde den Markt verlässt. Für diese Anwendungen werden weitaus leistungsstärkere Modelle als die hier vorgestellten benötigt. Es wird jedoch deutlich, welches hohe Potenzial der Einsatz von KI zur Erkennung und Klassifizierung von Objekten im Lebensmittel-Einzelhandel bietet.

Literaturverzeichnis

Apúntateuna. (05. Mai 2023). Abgerufen am 02. August 2023 von https://www.apuntateuna.es/deutschland/wie-viele-supermarkte-gibt-es-in-deutschland.html

Chollet, F. (2018). *Deep Learning mit Python und Keras.* (K. Lorenzen, Übers.) Frechen: mitp Verlags GmbH & Co. KG.

Dittrich, R. (22. Juli 2023). *Merkur*. Abgerufen am 02. August 2023 von https://www.merkur.de/verbraucher/discounter-smart-stores-einkaufen-shoppen-edeka-rewe-aldi-technik-ki-supermarkt-zr-92415757.html

Oldenthal, J. (11. Juni 2019). *mi.uni-koeln.de.* Abgerufen am 28. Oktober 2022 von http://www.mi.uni-koeln.de/wp-znikolic/wp-content/uploads/2019/06/11-Odenthal.pdf

Oltean, M. (2017). *Kaggle*. Abgerufen am 17. Juli 2023 von https://www.kaggle.com/datasets/moltean/fruits?resource=download

Tensor Flow Hub. (22. Oktober 2022). Von https://tfhub.dev/google/imagenet/efficientnet_v2_imagenet21k_ft1k_b0/feature_vector/2 abgerufen

V. Le, Q., & Tan, M. (23. Juni 2021). *arxiv.org*. Abgerufen am 25. Oktober 2022 von https://arxiv.org/abs/2104.00298

Anhang:

Jupyter Notebook 1: EDA_and_CNN_Fruits-Dataset.ipynb

Jupyter Notebook 2: Transfer Learning_Inference_Fruits-Dataset.ipynb