KI-basiertes Computer-Vision-System zur Qualitäts- und Größenbestimmung von Kartoffeln

Andreas Schliebitz[1], Henri Graf[1], Tobias Wamhof[1], Heiko Tapken[1] und Andreas Gertzen[2]

Abstract: Diese Arbeit untersucht die Weiterentwicklung einer stichprobenbasierten zu einer kontinuierlichen Qualitätsmessung von Kartoffellieferungen. Das dafür entwickelte KI-basierte Computer-Vision-System lokalisiert mithilfe eines YOLOv5-Detektors Kartoffeln auf einem Förderband mit einer Genauigkeit von 0,96 mAP@[.5:.95]. Eine anschließende Qualitätsbestimmung der detektierten Kartoffeln erfolgt mit einem EfficientNetV2-Klassifkator, der zur Familie der Convolutional Neural Networks zählt. Dieser zeigt auf einem qualitativ hochwertigen Referenzdatensatz eine Genauigkeit von 96 % auf acht Mängelklassen, welche auf dem zu erweiternden Förderband-Datensatz bei zwei Klassen auf 81 % und bei drei Klassen auf 72 % abfällt. Das Quadratmaß, Volumen und Gewicht einer Kartoffel werden über Segmentierungsmasken und Tiefenbilder approximiert. Zur echtzeitfähigen Annäherung der Geometrie wird anhand dieser Daten für jede erkannte Kartoffel ein triaxialer Ellipsoid berechnet. Weiterhin wird ein Ansatz zur Verbesserung der mit einem optimalen Schwellenwertalgorithmus berechneten Segmentierungsmasken auf Basis eines Mask R-CNN Segmentierungsmodells erarbeitet.

Keywords: Computer Vision, Kartoffel, Qualitätsmanagement, Künstliche Intelligenz, Agri-Food

1 Einleitung

Im Rahmen des Forschungsprojekts Agri-Gaia wird in Zusammenarbeit mit der Wernsing Feinkost GmbH untersucht, ob eine echtzeitfähige Qualitäts- und Größenbestimmung von Kartoffeln mit einem KI-basierten Computer-Vision-System umgesetzt werden kann. Ein solches Systems ist notwendig, da sowohl die Qualitäts- als auch Größenverteilung einer Kartoffellieferung aktuell an einer häufig nicht repräsentativen Stichprobe abgeschätzt wird. Der in dieser Näherung enthaltene Schätzfehler erlaubt in den meisten Fällen keine präzise Steuerung nachgelagerter Produktionsschritte, die aus wirtschaftlichen Gründen eine maximale Verwertung der Kartoffel anstreben. Durch eine genauere Erfassung der Qualitäten und Größen kann der Warenausschuss während der Produktion verringert und es können mehr Lebensmittel aus derselben Menge an Kartoffeln hergestellt werden.

Der Einsatzort des in dieser Arbeit vorgestellten Computer-Vision-Systems befindet sich unmittelbar nach einer Waschstraße, welche die angelieferten Kartoffeln von Verschmutzungen befreit. Die zu diesem Zeitpunkt sichtbar gewordenen Mängel auf der Kartoffeloberfläche werden von einem planparallel zum Vereinzelungsband

[1] HS Osnabrück, Fakultät IuI, Albrechtstr. 30, 49076 Osnabrück, a.schliebitz@hs-osnabrueck.de, h.graf@hs-osnabrueck.de, t.wamhof@hs-osnabrueck.de, h.tapken@hs-osnabrueck.de

[2] Wernsing Feinkost GmbH, Kartoffelweg 1, 49632 Addrup-Essen/Oldb. andreas.gertzen@wernsing.de

ausgerichteten Kamera-Array erfasst und für die Erstellung eines gelabelten Bilddatensatzes gespeichert. Während des manuellen Labeling-Prozesses werden jeder aufgenommenen Kartoffel mindestens eine von elf Qualitätsklassen und genau eine Bounding-Box zur Lokalisierung zugewiesen. Auf Grundlage dieses annotierten Datensatzes werden im Laufe dieser Arbeit verschiedene KI-Modelle trainiert, welche sowohl die Erweiterung des Bilddatensatzes erleichtern als auch letztendlich zwischen Qualitäts- und Größenklassen unterscheiden können. Die vorliegende Arbeit umfasst im Anschluss an diese Einleitung einen Überblick über den aktuellen Stand der Forschung, gefolgt von den verwendeten Methoden, einer Quantifizierung der erzielten Ergebnisse sowie ihrer Diskussion und einem abschließenden thematischen Ausblick.

2 Stand der Forschung

In der Fachliteratur existieren aufgrund der weltweiten Funktion der Kartoffel als Grundnahrungsmittel eine Vielzahl von Veröffentlichungen, die sich sowohl mit der Qualitäts- als auch Geometriebestimmung dieser Feldfrucht beschäftigen. So entwickeln beispielsweise Su et al. [Su20] ein automatisches Kartoffelsortiersystem, das in einem nicht industriellen Kontext primär mit Tiefenbildern aus einer aktiven Tiefenkamera arbeitet. Die aufgezeichneten Kartoffeln werden mithilfe eines Softmax-Regressionsmodells und eines Convolutional Neural Networks (CNN) in sechs Qualitätsklassen unterteilt. Dieselbe Autorengruppe untersucht in einer vorangegangenen Veröffentlichung [Su18] die Möglichkeit der Gewichtsbestimmung unterschiedlich geformter Kartoffeln anhand von Tiefendaten und den daraus errechneten Volumina. Die aufgezeichneten Tiefenbilder von 110 Kartoffeln werden außerdem genutzt, um ein Virtual-Reality-System zu konzipieren, welches die händische Untersuchung von Kartoffeln in einem 3D-Raum simuliert.

Im Gegensatz dazu fokussieren sich Oppenheim et al. [Op19] auf die Klassifizierung von Mängeln auf der Kartoffeloberfläche. Der 400 verschiedene Kartoffeln umfassende Datensatz besteht aus Farbbildern, die während des Labelings über Bounding-Boxen auf die mängelbehafteten Areale der Kartoffelschale zugeschnitten werden. Als Klassifikator kommt erneut ein CNN zum Einsatz, das auf diesem Datensatz eine Genauigkeit von über 92 % erzielt. Die reine Klassifikation äußerer Mängel wird sowohl von Hasan et al. [Ha21] als auch Wang et al. [WX21] mithilfe verschiedener CNN-Architekturen durchgeführt. Als Datengrundlage kommen Farbbilder ganzer Kartoffeln zum Einsatz. Wang et al. verwenden in ihrem Deep-Transfer-Learning-Ansatz ein Region-based Fully Convolutional Network (R-FCN) basierend auf der ResNet101-Architektur. Sie erzielen damit eine maximale Klassifikationsgenauigkeit von 98,7 % auf drei Mängelklassen. Beide Veröffentlichungen nutzen pro Mängelklasse einige hundert bis wenige tausend gelabelte Exemplare, wobei Hasan et al. die Größe ihres Datensatzes durch Augmentierung verfünffachen.

Eine dieser Arbeit ähnelnde Veröffentlichung stammt von Pandey et al. [PKP18], in der auf Basis eines U-Net Segmentierungsmodells und verschiedenen Bildverarbeitungs-

algorithmen einzelne Kartoffeln auf einem Förderband lokalisiert werden. Die Geometriebestimmung wird ausschließlich auf Grundlage der erzeugten Segmentierungsmasken in der 2D-Ebene des Förderbands durchgeführt. Für die Geometriebestimmung werden keine Tiefendaten erhoben. Die Klassifikation von sichtbaren Mängeln wird über ein Transfer-Learning-Ansatz realisiert. Abgrenzend zu diesen Veröffentlichungen wird in der vorliegenden Arbeit ein Prototyp eines echtzeitfähigen Systems zur Qualitäts- und Größenbestimmung von Kartoffeln auf einem Förderband vorgestellt. Die Größen-, Volumen- und Gewichtsbestimmung wird im Gegensatz zu Pandey et al. über Tiefenkameras um eine dritte Dimension erweitert. Eine Umsetzung erfolgt zusammen mit der Mängelklassifikation echtzeitfähig in Hard- und Software.

3 Methoden

Die Planung und Umsetzung des Computer-Vision-Systems umfasst im Wesentlichen die beiden Bereiche der Hard- und Software. Während mit der Kamerahardware Bilddaten sowohl für das Training als auch für die Inferenz von KI-Modellen aufgenommen werden, übernimmt die Software die Automatisierung zahlreicher Prozesse. Dazu gehört z. B. die Aktualisierung der Lokalisierungs-, Klassifikations- und Segmentierungsdatensätze mit neu gelabelten Kartoffeln sowie das Training-, Testing und Deployment von KI-Modellen. Der Lokalisierungsdatensatz wird für das Training eines Objektdetektors genutzt, der in einem ersten Schritt Kartoffeln auf einem Vereinzelungsband über Bounding-Boxen lokalisiert. Nach der Lokalisierung werden in einem händischen Labeling-Prozess die Bounding-Boxen mit Qualitätsklassen versehen. Der so gebildete Klassifikationsdatensatz besteht aus einer Menge von Qualitätsklassen mit einer jeweils variablen Anzahl an Bildern von Kartoffeln dieser Qualität. Das Ziel des mit diesem Datensatz trainierten Klassifikators ist die Unterscheidung zwischen Kartoffeln unterschiedlicher Qualitätsklassen. Basierend auf dem Lokalisierungsdatensatz wird ein Segmentierungsdatensatz erstellt, der für das Training eines Segmentierungsmodells genutzt wird. Dieses soll die Genauigkeit der mittels optimalem Schwellenwertalgorithmus erstellten Binärmasken verbessern und somit die Genauigkeit der Geometriebestimmung erhöhen. Eine Zusammenfassung dieses Ablaufs ist in Abbildung 1 dargestellt.

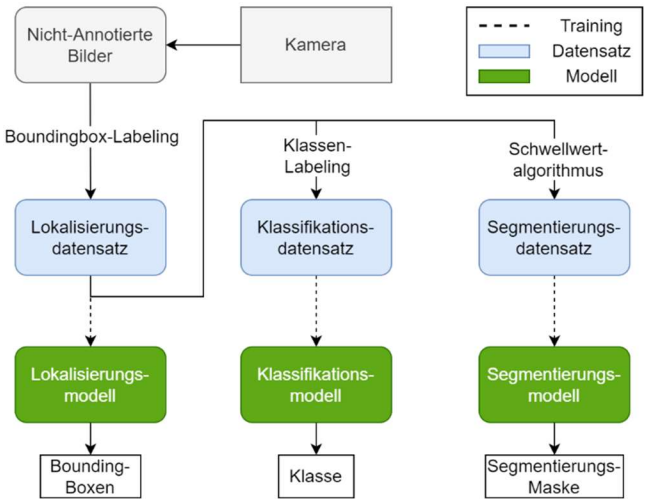

Abb. 1: Ablauf der KI-basierten Qualitäts- und Geometriebestimmung von Kartoffel

3.1 Kameraaufbau

Der Kameraaufbau wird über einem zwei Meter breiten Vereinzelungsband in einer Höhe von 550 mm installiert. Im Gegensatz zu herkömmlichen Förderbändern ermöglicht ein Vereinzelungsband die räumliche Trennung einzelner Kartoffeln. Als Kameras kommen drei OAK-D PoE der Firma Luxonis zum Einsatz, welche in einer Aluminiumkonstruktion mit Streulichtblende eingefasst sind. Die Kameras blicken orthogonal von oben auf das Vereinzelungsband und weisen zueinander einen Abstand von 650 mm auf (s. Abb. 2).

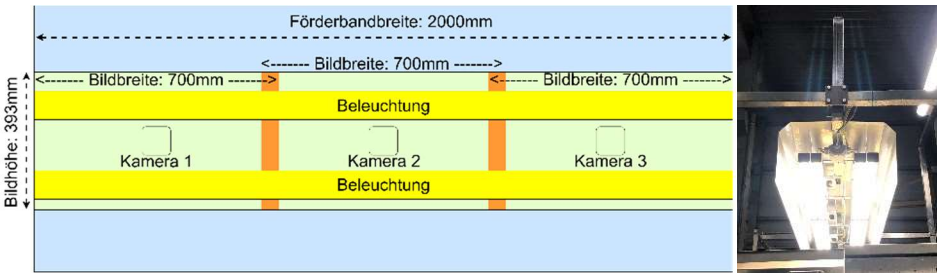

Abb. 2: Kameraaufbau über Vereinzelungsband bestehend aus drei RGBD-Kameras

Jede dieser Kameras nimmt zeitgleich Farb- und Tiefenbilder auf. Die Auflösung der Farbbilder (RGB) beträgt 1920×1080 Pixel, wohingegen die Tiefenbilder für eine passgenaue Überlagerung mit den Farbbildern von ihrer nativen Auflösung (1280×720 Pixel) auf 1920×1080 Pixel hochskaliert werden. Der Datenaustausch und die Stromversorgung der Kameras wird über einen Gigabit-Ethernet-Switch mit PoE-

Funktionalität (IEEE 802.3) abgewickelt. Die drei Kameras sind in einem lokalen Netzwerk über Cat6-Ethernet-Kabel mit einem Edge-Rechner verbunden. Dieser Computer besitzt neben einem Intel i7-6700 Vierkernprozessor, 500 GB NVMe Fest- und 32 GB Arbeitsspeicher eine NVIDIA Quadro P6000 Grafikkarte zur hardware-beschleunigten Ausführung von KI-Modellen.

Die Beleuchtung besteht aus insgesamt acht LED-Leuchtstoffröhren, die in zwei Gruppen zu jeweils einer Balkenbeleuchtung zusammengeschaltet werden. Durch eine gute Ausleuchtung des Förderbands können die RGB-Kameras mit einer kurzen Verschlusszeit von 450 μs und einer Lichtempfindlichkeit (ISO) von 650 betrieben werden. Die resultierenden Aufnahmen sind, wie in Abbildung 3 zu sehen, kontrastreich und scharf.

Abb. 3: Aufnahmen der drei Kameras vom Vereinzelungsband mit Kartoffeln

Damit während der Inferenz alle Kartoffeln von den Kameras erfasst werden können, darf das System eine minimale Bildrate (FPS_{min}) von ungefähr $3{,}18\,\mathrm{s}^{-1}$ nicht unterschreiten. Diese berechnet sich als der Quotient aus der maximalen Förderbandgeschwindigkeit $v_{max} \approx 1{,}25\ \mathrm{m/s}$ und der realen Bildhöhe $h_{Welt} \approx 0{,}393\ \mathrm{m}$.

3.2 Lokalisierung

Für die Lokalisierung einzelner Kartoffeln wird in einem Transfer-Learning-Ansatz ein auf dem COCO-Datensatz [Li14] vortrainierter YOLOv5-Objektdetektor des Größentyps X6 verwendet. Dieser wird zunächst mit semi-synthetischen Trainingsdaten trainiert, um das spätere Labeling eines echten Kartoffeldatensatzes durch automatisches Einzeichnen von Bounding-Boxen zu erleichtern. Das rechts in Abbildung 4 dargestellte Beispiel einer solchen Trainingseingabe wird mithilfe eines Referenzdatensatzes und Aufnahmen des leeren Förderbands zufällig generiert. Diese semi-synthetischen Bilddaten ermöglichen das Training eines ersten einsatzfähigen Kartoffel-Detektors, indem sie die in Abbildung 3 gezeigten Aufnahmen des realen Förderbands imitieren.

Abb. 4: Generierung semi-synthetischer Trainingsdaten zur Lokalisierung von Kartoffeln

Aufgrund der computergestützten Generierung dieser Bilddaten können die für das Training eines ersten Kartoffeldetektors benötigten Label in Form von Bounding-Boxen pixelgenau erzeugt werden. Das auf den synthetischen Daten trainierte KI-Modell zeigt auf den realen Bilddaten eine gute Generalisierungsfähigkeit mit vereinzelt ungenauen Bounding-Boxen und nicht detektierten Kartoffeln. Die nachfolgende Abbildung 5 zeigt die Verbesserung der Lokalisierungsgenauigkeit eines zweiten YOLOv5-Detektors, der mit den durch Menschen korrigierten Bounding-Boxen des ersten Detektors trainiert wird.

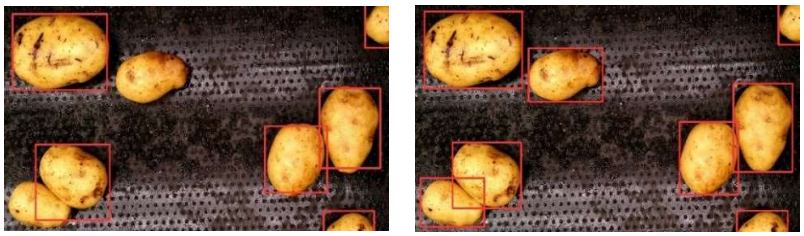

Abb. 5: Präzise Lokalisierung von Kartoffeln (rechts) nach Ausbesserung der synthetisch erzeugten Bounding-Boxen (links) und erneutem Training eines YOLOv5-Detektors

3.3 Geometriebestimmung

Für die echtzeitfähige Durchführung einer Geometriebestimmung wird das trainierte YOLOv5-Modell auf die Farbbilder des Förderbands angewendet. Die erzeugten Bounding-Boxen (rot) sind achsenorientiert und trennen einzelne Kartoffeln voneinander. Unabhängig von diesen Detektionen wird das Farbbild in einem nächsten Schritt in den HSV-Farbraum überführt, um mithilfe eines Hochpassfilters auf dem Sättigungskanal (S) störende Lichtreflexionen zu entfernen (s. Abb. 6). Eine Segmentierung der Kartoffeln erfolgt über eine optimale Schwellenwertbinarisierung des Wert-Kanals (V) nach dem Verfahren von Otsu [Ot79].

Abb. 6: Binarisierung mit dem Otsu-Verfahren nach Entfernung störender Reflexionen

Typischerweise verschmelzen bei einer Binarisierung manche Kartoffeln im Binärbild zu einem Objekt, welche mit herkömmlichen Methoden der Bildverarbeitung nur schwer wieder zu trennen wären. Die YOLOv5-Detektionen vermeiden dieses Problem im Binärbild, da die erzeugten Bounding-Boxen stets einzelne Objekte und somit auch ihre Vordergrundkomponenten voneinander separieren. Jede dieser Komponenten wird für die Längen- und Breitenbestimmung einer Kartoffel in ein Polygon umgewandelt (s. Abb. 7).

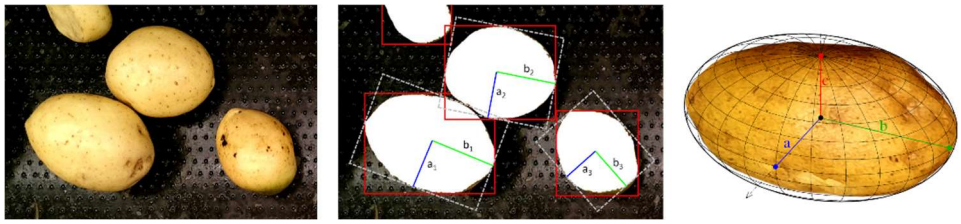

Abb. 7: Approximation der Kartoffelgeometrie mithilfe eines triaxialen Ellipsoids

Da einzelne Kartoffeln quer in ihrer axial ausgerichteten Bounding-Box liegen können, wird für eine genauere 2D-Geometriebestimmung ihr Minimum Rotated Rectangle (MRR, grau) berechnet. Ein Ansatz zur Annäherung des MRR bei Kartoffeln ist aufgrund ihrer Exzentrizität die Hauptkomponentenanalyse mit einer Rotation um den zwischen erster Hauptkomponente und x-Achse eingeschlossenen Winkel. Mit dem Rotating Calipers Algorithmus von Shamos [Sh78] existiert ein auf konvexen Polygonen schneller arbeitendes Verfahren, das in dieser Arbeit genutzt wird. Die Länge und Breite einer Kartoffel entspricht dann der Länge der kürzeren $(2a)$ und längeren Seite $(2b)$ des MRR in Millimetern.

Unter Miteinbeziehung des zeitgleich zum Farbbild aufgenommenen Tiefenbilds können anhand der beiden Halbachsen a, b sowie der halben Kartoffelhöhe c eine Volumen- und Gewichtsapproximation durchgeführt werden. Die Einträge eines Tiefenbilds D speichern die Entfernungen von der Kamera zu jedem Oberflächenpunkt in Millimetern. Das für die halbe Kartoffelhöhe c benötigte Höhenbild H berechnet sich über die Differenz eines Tiefenbilds des leeren Förderbands D_r (Referenz) und D über $H = D_r - D$.

Aufgrund der angestrebten Echtzeitfähigkeit des Systems wird das Volumen einer Kartoffel nicht über rechenintensive Punktwolken oder Dreiecksgitter, sondern über einen triaxialen Ellipsoid mit den drei Halbachsen a, b, c angenähert. Nach einer Abschätzung des Volumens über $V_{\text{Kartoffel}} \approx \frac{4}{3}\pi abc\,\text{mm}^3$ kann unter Annahme einer Kartoffeldichte[3] von $\rho \approx 1{,}0851\,\text{mg/mm}^3$ ihr Gewicht über $m_{\text{Kartoffel}} \approx V_{\text{Kartoffel}} \cdot \rho\,\text{mg}$ angenähert werden.

Die in diesem Abschnitt durchgeführte Binarisierung der Farbbilder funktioniert bei hellen Kartoffeln aufgrund des hohen Kontrasts zum dunklen Förderband besonders gut. Da dieser Kontrast bei dunkleren Kartoffeln abnimmt, können ihre Konturen durch die erzeugten Binärmasken nur weniger genau abgebildet werden. In diesen Fällen kann für eine genauere Konturdetektion ein KI-basiertes Segmentierungsverfahren eingesetzt werden. Aufgrund der Generalisierungsfähigkeit solcher Modelle besteht die Möglichkeit, dass ein auf den nicht einwandfreien Binärbildern trainiertes Segmentierungsmodell die Genauigkeit der Geometriebestimmung durch präzisere Freistellung der Kartoffeln verbessern kann.

[3] Durchschnittliche Dichte einer Kartoffel aus: https://blogs.imperial.ac.uk/physics-of-cooking/2011/03/09/potato-density-changes-with-age-karim-bahsoon (Stand: 09.12.2022)

3.4 Klassifikation

Zur Durchführung einer Mängelklassifikation werden mit dem in Abschnitt 3.1 beschriebenen Kameraaufbau Farbbilder von Kartoffeln aufgenommen und für ein händisches Labeling gespeichert. Der aus elf Mängelklassen bestehende Bilddatensatz wird mithilfe des Annotationsprogramms CVAT[4] (Computer Vision Annotation Tool) erstellt. Der Labeling-Prozess in CVAT besteht aus einer Korrektur fehlender oder ungenauer Bounding-Boxen sowie der Zuweisung mindestens einer der folgenden elf Qualitätsklassen: Wachstumsrisse (growth crack), mängelfrei (no defect), Trockenfäule (dryrot), Nassfäule (wet rot), Schorf (scab), mechanische Beschädigung (mechanical damage), grün (green), welk (withered), Wurmfraß (worm damage), Rüsselkäferbefall (weevilled) und Schwarzfleckigkeit (black spot).

Aufgrund des natürlichen Ursprungs von Kartoffeln entsteht nach dem Labeling-Prozess ein großes Ungleichgewicht zwischen den einzelnen Mängelklassen. Eine solche Unausgewogenheit kann während der Evaluierung eines Klassifikators die Leistungsmetriken Precision (P), Recall (R) und $F_{\beta=1}$ verfälschen. Aus diesem Grund wird vor dem Training eine Angleichung der Klassen durchgeführt, indem aus allen Klassen so viele Bilder ausgewählt werden, wie in der jeweils kleinsten Klasse enthalten sind.

Die in CVAT annotierten Bilddaten werden für das Training eines Convolutional Neural Networks verwendet. Die Kartoffeln werden anhand ihrer Bounding-Boxen aus den Förderbandbildern ausgeschnitten und über ihre erste Annotation in gleichnamige Verzeichnisse einsortiert. Für eine binäre Klassifikation werden alle mängelbehafteten Exemplare zu einer Klasse „defective“ zusammengefasst. Ein im Ergebniskapitel aufgeführtes Experiment umfasst die Klassifikation des Wernsing-Datensatzes, der in Abschnitt 3.2 für die Erzeugung der semi-synthetischen Förderbandbilder genutzt wird.

Als Klassifikator wird die von Torchvision (PyTorch Vision) bereitgestellte und auf dem ImageNet-1K Datensatz [Ru15] vortrainierte EfficientNetV2S-Architektur verwendet. Mit 21,5 Millionen Parametern ist diese CNN-Architektur kleiner, moderner und auf dem ImageNet-1K Datensatz um 8 % leistungsstärker als ein herkömmliches ResNet50. Das Training und die Auswertung des Klassifikators wird auf einzelnen Kartoffeln durchgeführt. Die Fähigkeit des YOLOv5-Detektors zwischen mehreren Klassen unterscheiden zu können, wird aufgrund des noch vergleichsweise kleinen Datensatzes nicht genutzt. Zur Verbesserung der Genauigkeit wird eine dedizierte Klassifikator-Architektur in einem zweistufigen System aus Lokalisierung (YOLOv5) und Klassifikation (EfficientNetV2S) eingesetzt.

3.5 Verwertung

Die anhand der Bounding-Boxen und Segmentierungsmasken bestimmten Geometrien, Klassen und Volumina werden zum Zeitpunkt der Detektion mit einem Zeitstempel

[4] Quelloffene Anwendung zur Annotation von Bild- und Videodaten: https://cvat.ai (Stand: 09.12.2022)

versehen. Die gesammelten Daten werden im JSON-Format über einen Publish-Subscribe-Mechanismus kontinuierlich bereitgestellt. Es werden keine Nachrichten gepublisht, wenn keine Kartoffeln erkannt werden. An das System per MQTT angebundene Applikationen (Datensenken) können die Detektionen empfangen und beliebig weiterverarbeiten. Die Speicherung der Daten erfolgt in einer relationalen PostgreSQL-Datenbank, welche die eintreffenden Detektionen in Blöcken von 1000 Nachrichten empfängt. Auf diese Weise wird eine echtzeitfähige Persistierung aller Detektionen gewährleistet.

Aus den gesammelten Daten können verschiedene Berichte generiert werden. Die Ergebnisse dieser Auswertungen werden in einem Fünfsekundentakt auf einem Grafana-Dashboard[5] visualisiert. Durch diese kontinuierliche Darstellung wird eine frühzeitige Reaktion auf unerwartete Qualitätsschwankungen im Warenstrom ermöglicht. Beispielhaft umgesetzte Auswertungen umfassen sowohl die Größen- und Massenverteilung sowie die absolute Anzahl an Kartoffeln einer Mängelklasse.

Durch Eingrenzen der Zeitstempel ist es zudem möglich, nur die Detektionen eines bestimmten Zeitraums auszuwerten. Mit dieser Funktion können rückwirkend Statistiken für die Zusammensetzung beliebiger Lieferungen erstellt und analysiert werden. Das System verarbeitet im normalen Betrieb etwa 200.000 Kartoffeln pro Stunde. Ein Datensatz hat in persistierter Form eine Größe von etwa 60 Bytes, sodass an einem Tag mit durchschnittlich 280-300 MB an Daten zu rechnen ist.

4 Ergebnisse

Die in diesem Kapitel vorgestellten Ergebnisse basieren auf einem Trainings- zu Testdatensatzverhältnis von 8:2. Trainings- und Testdatensätze werden stets derselben Grundgesamtheit entnommen und nicht mit andersartig erzeugten Datensätzen vermischt. Sämtliche Metriken, die in diesem Kapitel tabellarisch aufgelistet werden, beziehen sich auf den Testdatensatz. Das Training der KI-Modelle findet parallel auf vier Tesla V100 Grafikkarten einer NVIDIA DGX Station statt.

4.1 Lokalisierung

Die beiden Lokalisierungsdatensätze, die für das Training der in Tabelle 1 aufgelisteten YOLOv5-Detektoren verwendet werden, weisen zum einen generierte und zum anderen händisch ausgebesserte Bounding-Boxen auf. Der generierte Datensatz umfasst 10.000 Bilder mit einer Auflösung von 1920×1080 Pixeln und jeweils 18-24 Kartoffeln pro Bild. Die Gesamtsumme an Kartoffeln beläuft sich auf 230.484. Der gelabelte Förderband-Datensatz ist mit einer Größe von 22.331 Kartoffeln etwa zehnmal so klein.

[5] Quelloffene Anwendung zur visuellen Aufbereitung von Daten: https://grafana.com/ (Stand: 09.12.2022)

Label	Epochen	Batch	Eingabegröße	Box Loss	mAP@[.5:.95]
generiert	50	48	(720, 1280, 3)	0,0018	0,97
gelabelt	100	48	(720, 1280, 3)	0,0057	0,96

Tab. 1: Lokalisierungsgenauigkeiten des YOLOv5-Detektors auf generierten und echten Daten

Die beiden YOLOv5-Detektoren werden jeweils mit einer Batchgröße von 48 trainiert, wobei das zweite Modell mit dem gelabelten Datensatz doppelt so lange trainiert wird. Beide Modelle zeigen auf ihren jeweiligen Testdatensätzen unabhängig von der Kartoffelgröße sehr gute Lokalisierungsgenauigkeiten von über 0,95 mAP@[.5:.95] bei einem Box Loss nahe null. Der auf den gelabelten Daten trainierte Detektor wird für die Erweiterung des Klassifikations- und Segmentierungsdatensatzes verwendet.

4.2 Klassifikation

Die Untersuchung der Klassifizierbarkeit von Kartoffelmängeln wird mithilfe von zwei verschiedenen RGB-Datensätzen durchgeführt. Der erste dieser beiden Bilddatensätze wird von der Firma Wernsing bereitgestellt und umfasst acht Mängelklassen mit jeweils besonders repräsentativen Mängelausprägungen. Der zweite Datensatz beinhaltet die in elf Mängelklassen unterteilten Kartoffelbilder, welche mit dem Kamerasystem über dem Vereinzelungsband aufgenommen werden. Die Größen der jeweiligen Datensätze werden in der nachfolgenden Aufzählung durch Semikolons getrennt: 1. No defect (79; 13619), 2. Weevilled (1282; 10420), 3. Dryrot (1183; 2255), 4. Withered (0; 1725), 5. Mechanical damage (1210; 1527), 6. Scab (352; 1260), 7. Growth crack (0; 375), 8. Wet rot (124; 259), 9. Worm damage (0; 223), 10. Green (1250; 135), 11. Black spot (0; 108).

Eine Angleichung der Klassengrößen auf die jeweils kleinste ausgewählte Klasse findet nur bei den Klassifikationsdurchläufen des Förderband-Datensatzes statt. Alle EfficientNetV2S-Modelle werden in 50 Epochen mit dem AdamW-Optimierungsverfahren [LH18] und einer Verringerung der Lernrate um den Faktor zehn nach jeweils zehn Epochen trainiert. Der auf dem Wernsing-Datensatz trainierte Klassifikator liefert aufgrund der visuell eindeutigen Mängelausprägungen eine sehr gute Testgenauigkeit von über 96 % auf acht Qualitätsklassen (s. Tab. 3). Eine binäre Klassifikation des Förderband-Datensatzes zeigt eine Genauigkeit von etwa 81 %, die durch ein Hinzufügen der jeweils nächst größeren Mängelklasse aktuell noch auf bis zu 59 % bei vier Klassen abfällt.

Datensatz	Klassen	Batch	Acc@1	Acc@2	P	R	$F_{\beta=1}$
Wernsing	Alle	32	96,06%	99,94%	1,00	0,96	0,96
Förderband	1, Rest	32	80,55%	100,0%	0,99	0,81	0,89
Förderband	1, 2, 3	8	71,83%	91,37%	0,98	0,72	0,82
Förderband	2, 3, 4	8	71,13%	90,67%	0,99	0,71	0,81
Förderband	1, 2, 3, 4	8	59,08%	82,89%	0,96	0,59	0,70

Tab. 2: Erzielte Klassifikationsgenauigkeiten auf dem Wernsing- und Förderband-Datensatz

4.3 Segmentierung

Für eine Untersuchung, ob die mit dem Otsu-Verfahren erstellten Segmentierungsmasken durch ein KI-basiertes Segmentierungsmodell verbessert werden können, wird der aus 1980 Förderbandbildern bestehende Lokalisierungsdatensatz binarisiert. Die Binärmasken werden zusammen mit den Bounding-Boxen für das Training eines Mask R-CNN Modells genutzt. Dieses wird in 50 Epochen auf einer Klasse unter Verwendung des AdamW-Optimierers und einer Batchgröße von zwei trainiert. Das trainierte Modell zeigt auf den Testdaten sowohl eine hohe Segmentierungs- als auch Bounding-Box-Genauigkeit von 0,94 bzw. 0,93.

5 Diskussion und Ausblick

Das in dieser Arbeit präsentierte System zur Bestimmung von Kartoffelqualitäten und Größen ist in der Lage, sehr gute Genauigkeiten in der Lokalisierung und erste gute Ergebnisse in der Segmentierung einzelner Kartoffeln zu erreichen. Letzteres hat direkten Einfluss auf die Geometriebestimmung, weshalb ein KI-basierter Segmentierungsansatz in Zukunft noch genauer zu untersuchen ist. Eine echtzeitfähige Umsetzung der Detektion und Klassifikation ist in Abhängigkeit der verwendeten Edge-Hardware und Netzgröße möglich. Die Inferenzrate kann dabei über eine Abstimmung der KI-Modelle auf die zur Verfügung stehende Edge-Hardware gesteuert werden. Für eine Verbesserung der Klassifikationsergebnisse wird ein größerer Datensatz mit aussagekräftigeren Exemplaren pro Mängelklasse benötigt. Der Klassifikator tendiert bei mehr als zwei bis drei Klassen dazu, diese mit anderen ähnlichen Klassen zu verwechseln (bspw. „no defect“ mit „weevilled“). Grund hierfür ist wahrscheinlich die zum Teil noch unzureichende Mängelausprägung in den jeweiligen Bilddatensätzen, gepaart mit einer vergleichsweise kleinen Grundgesamtheit. Eine Erweiterung des Förderband-Datensatzes wird durch den sehr genau arbeitenden Detektor erleichtert. Der bisherige Datensatz könnte im Rahmen eines Self-Supervised-Learning-Ansatzes zur Feinabstimmung eines selbstlernenden Klassifikators genutzt werden. KI-Modelle dieser Art benötigen während des Trainings keine Label, sind aber extrem groß und daher langsam zu trainieren.

Die mit dem System aufgezeichneten Kartoffelgrößen und Massen erscheinen, gemessen an ihren Absolutwerten und Normalverteilungen, plausibel. Eine genauere Annäherung der Kartoffelgeometrie ist vermutlich mit aktiven Tiefenkameras und Punktwolken möglich. Die Echtzeitfähigkeit eines solchen Systems wird jedoch durch den erhöhten Rechenaufwand wahrscheinlich reduziert. Durch eine Einbettung dieses KI-Systems in bestehende Prozesse der Firma Wernsing wird zukünftig mit einer verbesserten Rohstoffverwertung gerechnet. Das entstandene Computer-Vision-System birgt aufgrund seiner Modularität und Übertragbarkeit auf andere landwirtschaftliche Erzeugnisse ein großes Wertschöpfungspotenzial.

Förderhinweis: Wir danken dem Bundesministerium für Wirtschaft und Klimaschutz für die Förderung des Projektes Agri-Gaia unter dem Förderkennzeichen 01MK21004G.

Literaturverzeichnis

[Ha21] Hasan, Md; Zahan, Nusrat; Zeba, Nahid; Khatun, Amina; Haque, Mohammad Reduanul et al.: A Deep Learning-Based Approach for Potato Disease Classification. In: Computer Vision and Machine Learning in Agriculture, S. 113-126. Springer, 2021.

[LH18] Loshchilov, Ilya; Hutter, Frank: Decoupled Weight Decay Regularization. In: International Conference on Learning Representations. 2018.

[Li14] Lin, Tsung-Yi; Maire, Michael; Belongie, Serge; Hays, James; Perona, Pietro; Ramanan, Deva; Dollár, Piotr; Zitnick, C Lawrence: Microsoft coco: Common objects in context. In: European conference on computer vision. Springer, S. 740-755, 2014.

[Op19] Oppenheim, Dor; Shani, Guy; Erlich, Orly; Tsror, Leah: Using deep learning for image-based potato tuber disease detection. Phytopathology, 109(6):1083-1087, 2019.

[Ot79] Otsu, Nobuyuki: A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics, 9(1):62-66, 1979.

[PKP18] Pandey, Nikhil; Kumar, Suraj; Pandey, Raksha: Grading and defect detection in potatoes using deep learning. In: International conference on communication, networks and computing. Springer, S. 329-339, 2018.

[Ru15] Russakovsky, Olga; Deng, Jia; Su, Hao; Krause, Jonathan; Satheesh, Sanjeev; Ma, Sean; Huang, Zhiheng; Karpathy, Andrej; Khosla, Aditya; Bernstein, Michael et al.: Imagenet largescale visual recognition challenge. International journal of computer vision, 115(3):211-252, 2015.

[Sh78] Shamos, Michael Ian: Computational geometry. Yale University, 1978.

[Su18] Su, Qinghua; Kondo, Naoshi; Li, Minzan; Sun, Hong; Al Riza, Dimas Firmanda; Habaragamuwa, Harshana: Potato quality grading based on machine vision and 3D shape analysis. Computers and electronics in agriculture, 152:261-268, 2018.

[Su20] Su, Qinghua; Kondo, Naoshi; Al Riza, Dimas Firmanda; Habaragamuwa, Harshana: Potato quality grading based on depth imaging and convolutional neural network. Journal of Food Quality, 2020, 2020.

[WX21] Wang, Chenglong; Xiao, Zhifeng: Potato Surface Defect Detection Based on Deep Transfer Learning. Agriculture, 11(9):863, 2021.