

Unravel – Bilderkennung mit neuronalen Netzen

Gräßl, C.
Donau-Gymnasium Kelheim

DOI: 10.18420/ibis-01-01-09

Zusammenfassung

Die zunehmende Bedeutung von Künstlicher Intelligenz im Alltag hat dazu geführt, dass die Kultusministerien verschiedener Bundesländer dieses Thema in Lehrplänen allgemeinbildender Schulen verankert haben. Da dieser Lernbereich neu ist, müssen auch innovative didaktische Ansätze entwickelt werden. Dies stellt aufgrund weniger Erfahrungswerte eine besondere Herausforderung dar. In dieser Arbeit wird ein praxisorientiertes Konzept vorgestellt, in dem Schülerinnen und Schüler ab der oberen Mittelstufe selbstständig alle Komponenten eines Bilderkennungssystems, angefangen von der Bildaufnahme bis hin zur Klassifizierung, explorativ erleben. Ziel ist es zum einen, einen grundlegenden Einblick in Klassifikation mit neuronalen Netzen zu vermitteln. Zum anderen sollen die Schülerinnen und Schüler verstehen, wie Systeme des überwachten Lernens allgemein aufgebaut sind, wie man sie bewertet und wo deren Grenzen liegen.

Einleitung

In vielen Supermärkten wird die Kundschaft gebeten, ihr Obst oder Gemüse selbst an der Waage zu etikettieren. An der Kasse geht es dann etwas schneller, da nur noch der Strichcode eingescannt werden muss. Kundinnen und Kunden müssen dabei aus oft fünfzig verschiedene Obst- und Gemüsesorten die richtige identifizieren. Um dies zu erleichtern haben einige Supermarktketten „intelligente“ Waagen eingeführt. Wird eine Frucht auf die Waagschale gelegt, erhält man einen Vorschlag, um welche Sorte es sich dabei höchstwahrscheinlich handelt, zusammen mit einigen weiteren, die ebenfalls in Frage kommen könnten. Doch woher weiß die Waage, dass genau diese Obstsorte gewogen werden soll?

Auf der Waage befindet sich eine Kamera und ein Computer, der mithilfe eines Algorithmus versucht, die Frucht zu klassifizieren. Genaugenommen handelt es sich hierbei um ein System zur Objekterkennung, das ein Verfahren des maschinellen Lernens und somit Künstliche Intelligenz verwendet. Natürlich müssen dem System zuvor für jede Obst- und Gemüsesorte viele Bilder bereitgestellt werden, damit es später auch die Fruchtsorten erkennen kann, die von den Kundinnen und Kunden auf die Waage gelegt werden können. Da hier das Training des Algorithmus mit bekannten (beschrifteten) Daten erfolgt, wird diese Vorgehensweise als überwachtes Lernen bezeichnet.

In dem hier vorgestellten Konzept lernen Schülerinnen und Schüler anhand des oben beschriebenen Szenarios, wie Bilderkennungssysteme grundsätzlich funktionieren. Die Jugendlichen bringen dazu Früchte mit in den Unterricht und erkunden explorativ mit Hilfe des didaktischen Tools *Unravel* alle Schritte, angefangen von der Bildaufnahme bis hin zur Klassifikation mithilfe eines neuronalen Netzes. Es handelt sich somit um einen Ansatz, der sich durch haptische Komponenten auszeichnet und von *KI-Unplugged*¹ (Seegerer/Lindner/Romeike 2019) inspiriert ist.

Bei *Unravel* (siehe Abbildung 1) handelt es sich um eine Webanwendung, die in JavaScript implementiert wurde und somit auf allen modernen Browsern und unabhängig vom Betriebssystem verwendet werden kann. Zum Start muss lediglich die URL <https://klassenkarte.de/unravel> geöffnet werden. Danach findet eine einmalige Kommunikation mit dem Server statt, um die Programmdateien herunterzuladen. Daten werden nur lokal und temporär auf dem Computer der Schülerinnen und Schüler gespeichert, was dieses Werkzeug aus Datenschutzperspektive unbedenklich macht.

Die Bildaufnahme der Früchte erfolgt über eine am Computer (in der Regel über USB) angeschlossene Kamera, die senkrecht nach unten gerichtet ist. In der Praxis haben sich hierbei Dokumentenkameras bewährt, die in den Schulen normalerweise in ausreichender Zahl vorhanden sind. Es ist nicht erforderlich, für jeden der Jugendlichen eine eigene Kamera bereitzustellen, da eine Gruppenarbeit empfehlenswert ist.

Während die Schülerinnen und Schüler in der ersten Phase selbstständig erforschen, wie bildverarbeitende Objekterkennung funktioniert, steht die Lehrkraft als „technischer Support“ zur Seite. Sobald die Jugendlichen nach ein bis zwei Stunden das System verstanden haben, wird

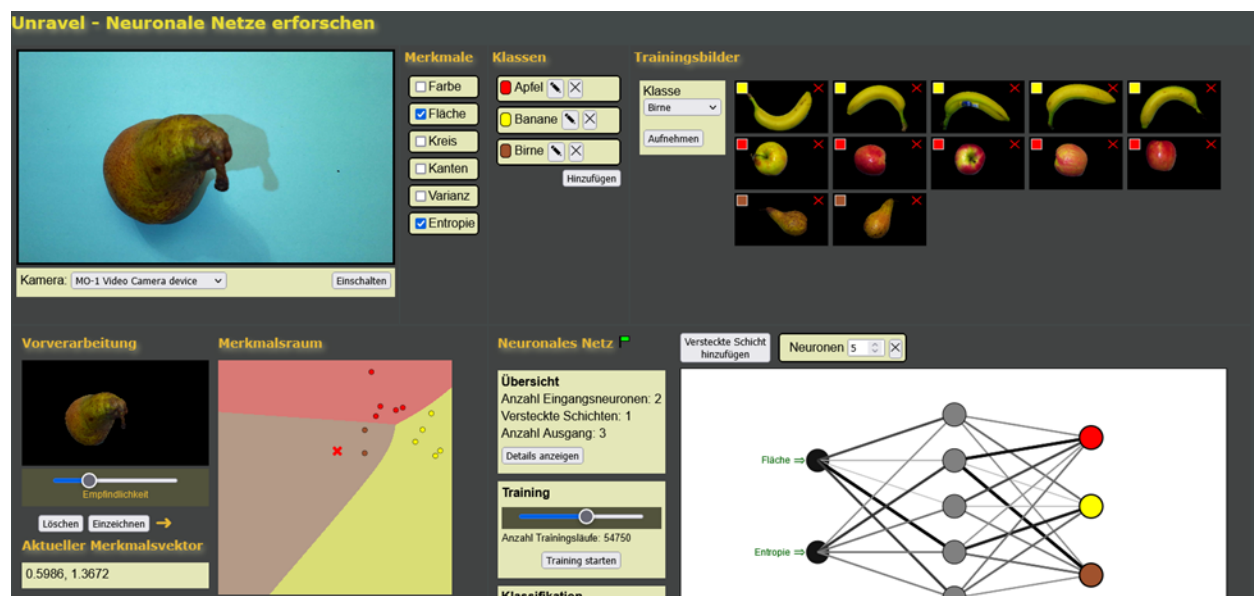


Abbildung 1: Benutzeroberfläche von Unravel

¹ <https://www.aiunplugged.org/>

das gewonnene Wissen strukturiert und gesichert. Abhängig von den individuell gesetzten Lernzielen der Lehrkraft können nun bestimmte Aspekte detaillierter betrachtet werden. Insbesondere um Künstliche Intelligenz zu entmystifizieren, können zumindest einige der verwendeten Merkmale des Systems (wie zum Beispiel die Größe der Frucht oder die Farbe) untersucht werden. Besonders die Auswirkungen verschiedener Konfigurationen des neuronalen Netzes (bezüglich der Anzahl und Struktur der Schichten) können analysiert und bewertet werden. Bei diesen Experimenten können wichtige Begriffe des maschinellen Lernens wie Validierungs- und Testdaten nicht nur eingeführt, sondern praktisch „erlebt“ werden.

Im nächsten Abschnitt wird der fachliche Hintergrund zu Bilderkennungssystemen anhand des Szenarios erläutert. Darauf folgt ein Vorschlag zur didaktischen Umsetzung. Der Artikel endet mit einem Fazit und einem Ausblick.

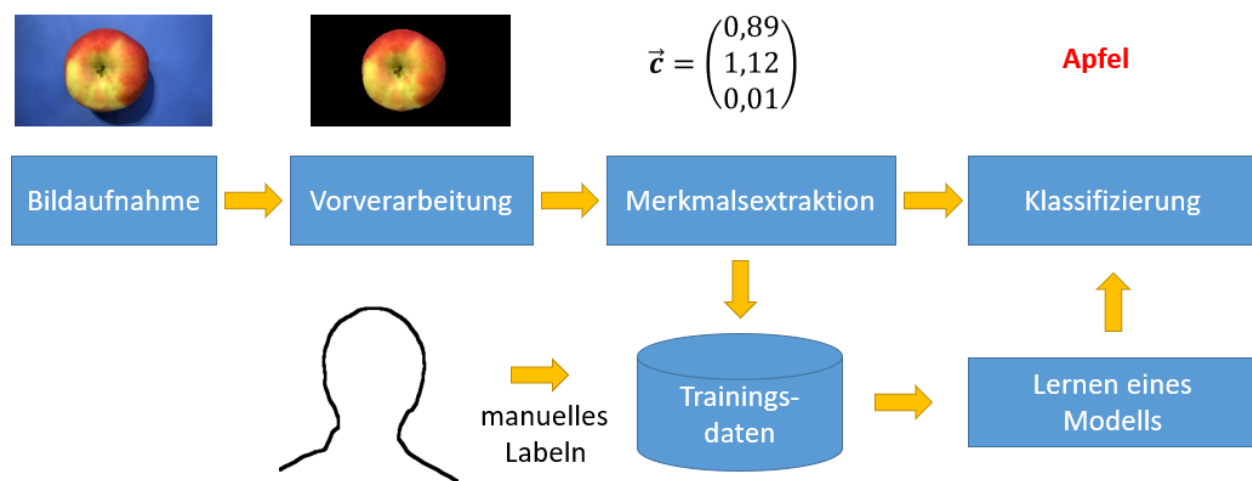


Abbildung 2: Aufbau eines Systems zur Klassifikation von Mustern (Niemann 1982), adaptiert auf „überwachtes Lernen“

Fachlicher Hintergrund

Obwohl der Begriff „Künstliche Intelligenz“ in den letzten Jahren vermehrt im täglichen Leben präsent ist, handelt es sich eigentlich um ein Thema mit einer langen Geschichte. Bereits vor über 40 Jahren wurden Systeme zur Klassifikation von Mustern erforscht. In Abbildung 2 ist der Aufbau eines solchen Systems im Überblick zu sehen (Niemann 1982), das in dieser Arbeit für ein Bilderkennungssystem im Rahmen des überwachten Lernens adaptiert wurde. Im Folgenden werden die Komponenten dieses Systems im Hinblick auf das Szenario der „Früchteerkennung“ beleuchtet.

Bildaufnahme

In der Regel werden Bilder mithilfe einer digitalen Kamera im Rastergrafikformat aufgenommen. Über die Zeilen- und Spaltennummer kann auf die Farbinformation eines Bildpunktes (Pixels) zugegriffen werden. Typischerweise werden drei Bytes pro Pixel verwendet, um die

Farbe anhand der Anteile von Rot, Grün und Blau zu repräsentieren. Je mehr Spalten und Zeilen ein Bild besitzt, desto höher ist auch seine Auflösung und desto mehr Details sind in den Bilddaten enthalten.

Vorverarbeitung

Die vorliegenden Bilddaten sind in der Regel noch nicht geeignet, um von Algorithmen verarbeitet zu werden. Beispielsweise durch die Anwendung von Filteroperationen werden die Qualität verbessert und Rauscheffekte unterdrückt. Zudem wird das zu erkennende Objekt vom Hintergrund getrennt (segmentiert), da die nachfolgenden Verfahren zur Merkmalsextraktion ansonsten weniger zuverlässig arbeiten oder sogar scheitern können. Im Fall von *Unravel* wird eine Segmentierung durchgeführt, indem zuerst der durchschnittliche Farbwert im Randbereich des Bildes berechnet wird. Dieser Wert repräsentiert dann die Farbe des Hintergrunds. Pixel, die sich von der Hintergrundfarbe in einem bestimmten (konfigurierbaren) Maß unterscheiden, werden als Punkte, die zum Objekt gehören, betrachtet. Es wird empfohlen, einen blauen Hintergrund zu verwenden, da in der Regel keine Frucht diese Farbe aufweist. Je nach Fruchtauswahl können auch andere Farben geeignet sein.

Merkmalsextraktion

Man verwendet den Begriff „Merkmal“, um messbare Eigenschaften oder Attribute zu beschreiben, die zur Identifikation von Objekten auf einem Bild verwendet werden können. Dabei werden verschiedene Merkmale wie beispielsweise geometrische Eigenschaften oder Farbinformationen genutzt. Es wird davon ausgegangen, dass Objekte derselben Klasse ähnliche Merkmale aufweisen, während Objekte unterschiedlicher Klassen auch in ihren Merkmalen differieren. Ein hilfreiches Merkmal, um Tomaten von Bananen auseinanderhalten zu können, ist beispielsweise der Rotanteil (0 für kein Rot, 1 für vollständig rot). Außerdem ist ein geometrisches Merkmal, das die Kreisförmigkeit des Objekts als Zahlenwert angibt, geeignet. Allerdings ist dieses Merkmal weniger geeignet, um einen Apfel von einer Tomate zu unterscheiden. Daher setzen Bilderkennungssysteme auf mehrere Merkmale zur Erkennung von Objekten. *Unravel* verwendet dabei unter anderem die folgenden Merkmale:

Flächenanteil

Da sich viele Früchte allein anhand ihrer Größe unterscheiden lassen, bietet es sich an, den Flächenanteil der Bildpunkte, die zur Frucht gehören, als Merkmal zu berechnen. Hierzu werden zunächst alle Bildpunkte gezählt, die zum Vordergrund, also zur Frucht, gehören. Anschließend wird dieser Wert durch die Gesamtzahl der Pixel im Bild dividiert, um den Anteil der Vordergrundfläche von dem gesamten Bild zu berechnen. Kleine Früchte wie Erdbeeren und Pflaumen haben somit im Bild einen kleineren Flächenanteil als Bananen oder Äpfel. Damit dieses Merkmal sinnvoll genutzt werden kann, muss die Kamera immer den gleichen Abstand zur Auflagefläche der Frucht sowie denselben Zoomfaktor haben. In Abbildung 3 wird die Berechnung des Flächenanteils für drei verschiedene Früchte veranschaulicht.

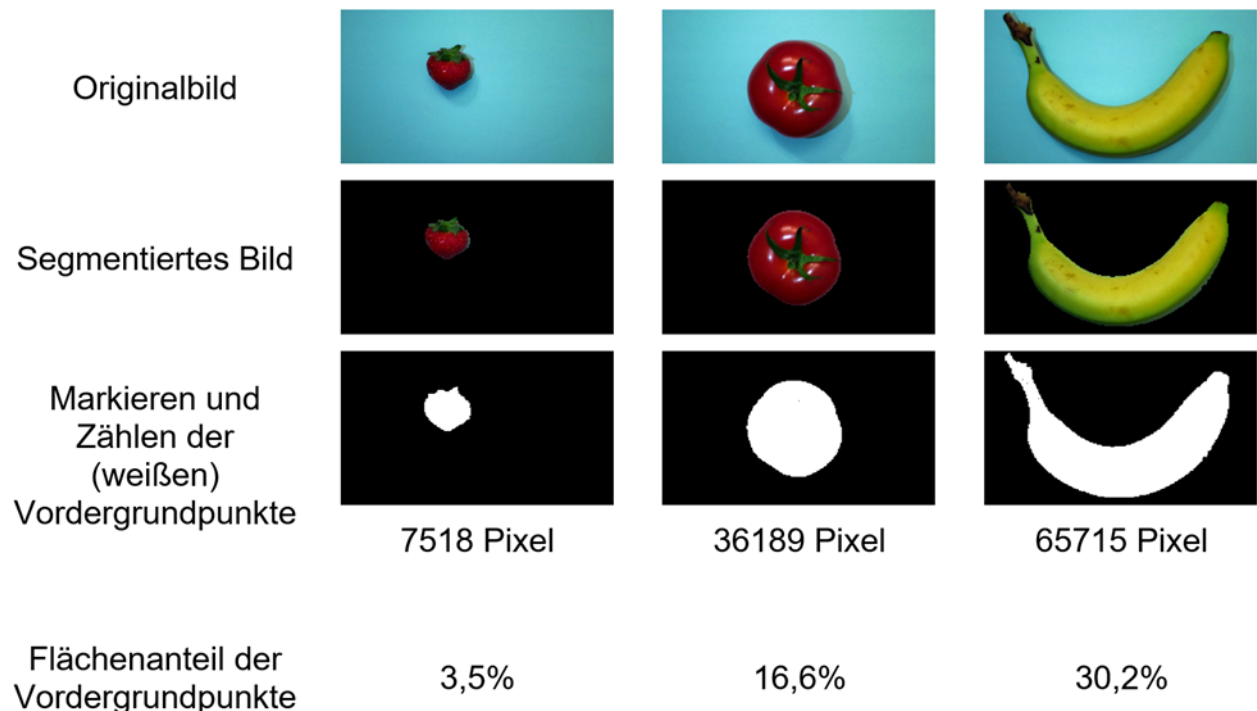


Abbildung 3: Der Flächenanteil als Merkmal für Früchte

Farbwinkel

Die Farbe einer Frucht stellt ein sehr charakteristisches Merkmal dar. Allerdings hat sich in der Praxis gezeigt, dass bereits geringe Schwankungen in der Beleuchtung ausreichen, um starke Abweichungen in den Rot-, Grün- und Blauanteilen zu verursachen. Aus diesem Grund ist es sinnvoll, die Farbe auf den helligkeitsunabhängigen Farbwinkel (engl. Hue) abzubilden. Dies geschieht durch eine einfache mathematische Berechnung. In Abbildung 4 sind Beispiele für den Farbwinkel verschiedener Früchte, normiert auf den Wertebereich von 0 bis 1, dargestellt.

Kreisförmigkeit

Äpfel und Tomaten erscheinen auf Bildern im Gegensatz zu Bananen oder Karotten fast kreisrund. Tatsächlich lässt sich die Kreisförmigkeit einer Figur mit einer Formel bestimmen, sofern der Umfang und der Flächeninhalt bekannt sind:

$$c = \frac{\text{Flächeninhalt}}{(\text{Umfang})^2}$$

Bei einem Kreis erhält man den Wert $c = \frac{\pi r^2}{(2\pi r)^2} \cong 0,0796$, was auch dem Maximalwert für eine Figur entspricht. Ein Quadrat hat einen Wert von $c = 0,0625$ und bei einem Rechteck mit einem Verhältnis der Seitenlängen von 4 zu 1 ist $c = 0,04$.

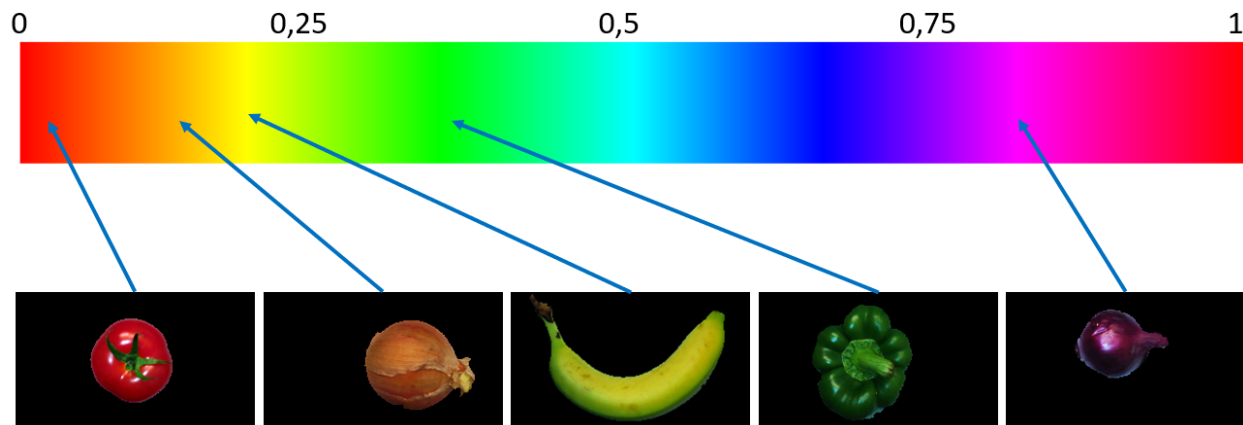


Abbildung 4: Der durchschnittliche Farbwinkel der Bildpunkte verschiedener Früchte normiert auf eine Skala von 0 bis 1

Der Flächeninhalt kann, wie beim Merkmal „Flächenanteil“ bestimmt werden, indem die Pixel gezählt werden, die dem Vordergrund zugeordnet sind. Den Umfang kann man berechnen, indem man alle Punkte zählt, die zum Objekt gehören und gleichzeitig an einem Punkt des Hintergrunds angrenzen.

Weitere Merkmale

Unravel bietet auch die Möglichkeit, die Merkmale *Entropie*, *Varianz* und *Kanten* zu nutzen. Ähnlich wie bei den zuvor betrachteten Merkmalen, können diese anhand der Farbanteile der Objektpunkte berechnet werden. Eine genauere Betrachtung würde aber den Rahmen dieser Arbeit sprengen. Daher wird auf einschlägige Literatur zur Bildverarbeitung wie zum Beispiel (Burger/Burge 2015) verwiesen. Auf der Projekthomepage² steht zudem Material für Schülerinnen und Schüler zur Verfügung, das als Orientierung dienen kann.

Trainingsdaten

Verfahren des überwachten Lernens benötigen Trainingsdaten, um ein Modell zu erstellen, das in der Lage ist, Vorhersagen (in unserem Fall Klassifizierungen) für neue Daten zu treffen. Bei *Unravel* bestehen die Trainingsdaten aus den Merkmalen von Bildern, die von der Person, die das Bild aufgenommen hat, mit den entsprechenden Klassenbezeichnern versehen (gelabelt) wurden. Generell kann man sagen, dass ein besseres Training des Modells möglich ist, je mehr repräsentative Trainingsdaten zur Verfügung stehen (Hastie/Tibshirani/Friedman 2009). Daher sollten auch im vorliegenden Fall für jede Fruchtart mehrere Trainingsdatensätze vorliegen.

Lernen eines Modells und Klassifizierung

Ein Modell kann als eine Art Funktion betrachtet werden, die Eingabedaten in Form von Merkmalen erhält und eine entsprechende Klasse berechnet. Solch ein Modell kann beispielsweise

² <https://klassenkarte.de/index.php/ki/unravel/>

auf einem Entscheidungsbaum basieren oder wie bei *Unravel* auf einer Perzeptron- bzw. neuronalen Netzwerkarchitektur. Eine gute Einführung zu diesen Modellen bietet die Handreichung für Informatik-Lehrkräfte an bayerischen Gymnasien (ISB 2023). Dieses Modell kann nach einer Trainingsphase zur Klassifizierung von Daten, in diesem Fall Bilder von Obst, verwendet werden.

In Abbildung 5 wird die Klassifizierung mittels eines neuronalen Netzes verdeutlicht. Zunächst wurde das Netz mit Trainingsbildern von drei verschiedenen Klassen (Apfel rot, Banane gelb und Kiwi braun) sowie den beiden Merkmalen *Flächenanteil* und *Kreisförmigkeit* trainiert. Anschließend wurde ein Bild einer weiteren Frucht aufgenommen, die nicht in den Trainingsdaten enthalten war und nun klassifiziert werden soll. Das Bild wird vom Programm segmentiert und die beiden Merkmale werden berechnet, um als Eingabe für das neuronale Netz zu dienen. Der Algorithmus ermittelt für jede bekannte Klasse einen Ausgabewert zwischen 0 und 1. Eine Ausgabe von 1 bedeutet, dass es sich höchstwahrscheinlich um die entsprechende Klasse handelt, während eine Ausgabe von 0 bedeutet, dass es sich mit ziemlicher Sicherheit nicht um die Klasse handelt. Das Ergebnis der Klassifizierung ist die Klasse mit dem höchsten Wert.

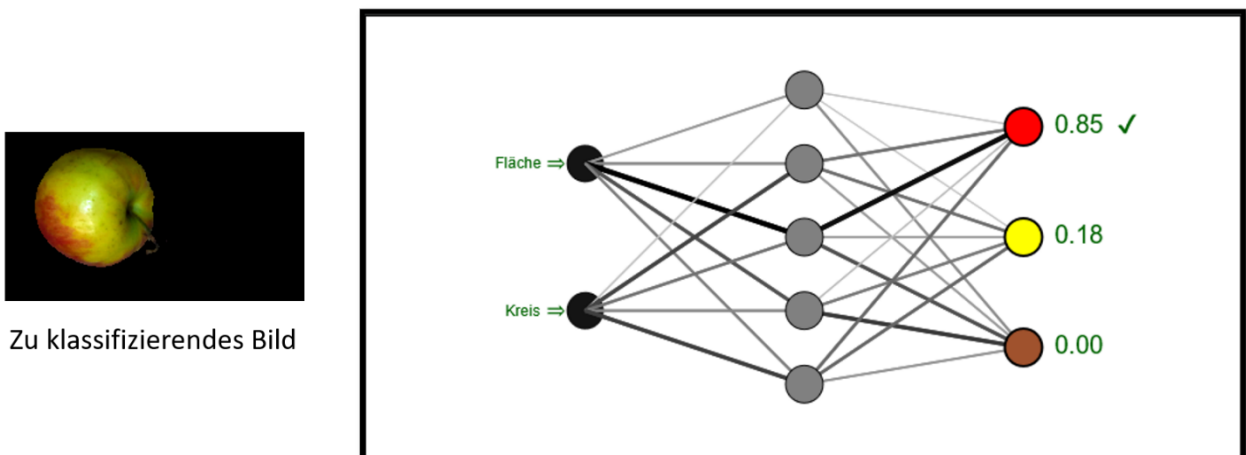


Abbildung 5: Das neuronale Netz klassifiziert das Bild als Apfel, da das entsprechende Ausgabeneuron (rot) den höchsten Wert liefert. Die Klasse Banane (gelbes Ausgabeneuron) erhält den zweithöchsten Wert und am wenigsten wahrscheinlich ist die Kiwi (braunes Ausgabeneuron).

Didaktische Umsetzung

Zu Beginn der Unterrichtssequenz bietet es sich an, den Schülerinnen und Schülern lediglich das Supermarkt-Szenario vorzustellen und ihnen den Arbeitsauftrag zu geben, mit Hilfe der Anwendung *Unravel* und der Kamera, dem Computer „beizubringen“, die mitgebrachten Früchte zu erkennen. Vorwissen bezüglich Künstlicher Intelligenz ist an dieser Stelle nicht erforderlich. Aufgrund des Aufbaus der selbsterklärenden Benutzeroberfläche sind die Schülerinnen und Schüler nach kurzer Zeit selbstständig in der Lage, den Systemaufbau zu

durchschauen. Als Lehrkraft sollte man in dieser Phase bereitstehen, um bei technischen Problemen zu unterstützen und Fragen zu beantworten. Die Jugendlichen erkennen dabei auch die Grenzen des Systems. Eine fleckige Banane wird häufig als Kiwi fehlklassifiziert oder ein roter Apfel als Tomate, falls nur mit grünen und gelben Äpfeln trainiert wurde. Zusätzlich werden natürlich nur Obstsorten erkannt, für die eine Klasse definiert wurde. Jedes Scheitern ist dabei als Gewinn zu sehen, da die Jugendlichen versuchen, die Ursachen zu identifizieren und Lösungen für ein besseres Ergebnis umzusetzen. Anschließend werden die aufgetretenen Probleme im Klassenverband gesammelt und damit grundsätzliche Grenzen derartiger Systeme erkannt.

Im nächsten Schritt empfiehlt es sich, in thementeiliger Gruppenarbeit, genauer auf die Merkmalsextraktion einzugehen. Jede Gruppe kann sich dabei mit einem der sechs vorhandenen Merkmale auseinandersetzen. Die Gruppeneinteilung kann hier leistungshomogen erfolgen. Das Merkmal *Flächenanteil* ist relativ leicht zu verstehen. *Farbwinkel* und *Kreisförmigkeit* sind etwas anspruchsvoller. Die drei weiteren Merkmale sind eher für leistungsstärkere Schülerinnen und Schüler geeignet³. In dieser Phase sollen die Jugendlichen auch experimentieren, um ihr zugeteiltes Merkmal zu erkunden. Das ist relativ einfach, da die Kamera ständig Bilder aufnimmt und die Merkmalspunkte in Echtzeit dargestellt werden.

Nachdem sich die Gruppen jeweils in „ihr“ Merkmal eingearbeitet haben, findet in einer zweiten Runde (beispielsweise durch Posterpräsentationen, „Kugellager“ oder „Gruppenpuzzle“) die Weitergabe der erarbeiteten Erkenntnisse statt. Je nach Klassensituation kann es günstig sein, sich auf wenige Merkmale zu beschränken, da es nicht zwingend nötig ist, dass jeder der Jugendlichen im Detail weiß, wie jedes Merkmal berechnet wird. Es geht vielmehr um das grundlegende Verständnis eines Verfahrens des maschinellen Lernens und die Entmystifizierung der Künstlichen Intelligenz. Am Ende der zweiten Phase sollte besprochen werden, dass sich Bilderkennungssysteme je nach Zielsetzung zwar in den Merkmalen unterscheiden können, die zugrundeliegende Idee jedoch erhalten bleibt. Parallelen können hier auch zur Gesichtserkennung gezogen werden. Typische Merkmale sind hierbei *Positionen* und *Größe* der Augen, Nase und Mund.

Durch den explorativen didaktischen Ansatz haben im Laufe der Sequenz schon viele Schülerinnen und Schüler entdeckt, dass sich das neuronale Netz weiter konfigurieren lässt. Auf die Frage, welche Einstellung, die beste ist, werden die Gruppen berichten, wo sie gute Erfahrungen gemacht haben, die sich jedoch sicherlich auch widersprechen. Je nach gewünschter fachlicher Tiefe können an dieser Stelle die Begriffe Test- und Validierungsdaten, aber auch Gütekriterien wie die „Genauigkeit“ thematisiert werden. Nach einer Einigung auf ein Setting (das heißt Aufteilung der Trainings- und Testdaten) können die Modelle mit den verschiedenen Konfigurationen und zugrundeliegenden Datensätzen bewertet werden.

³ Auf der [Projekthomepage](#) befindet sich hierzu Material, das von den Schülerinnen und Schülern verwendet werden kann.

Je nach Lehrplan beziehungsweise Zielsetzung der Sequenz könnte man hier tiefer in die Thematik der neuronalen Netze eintauchen oder auch andere Klassifikationsansätze wie den Entscheidungsbaum- oder den k-nächste-Nachbarn-Algorithmus behandeln. Würde man nach dem bayerischen Lehrplan der 11. Jahrgangsstufe unterrichten, so könnte man als nächstes das künstliche Neuron (Perzeptron) als Grundbaustein eines neuronalen Netzes thematisieren.

Was übrigens keinesfalls fehlen darf, egal, ob man das hier vorgestellte Konzept in den lehrplanmäßigen Unterricht integriert oder als separates Projekt durchführt, ist eine abschließende gesellschaftlich-kulturelle Betrachtung von Bilderkennungssystemen im Sinne der Dagstuhl-Erklärung (Brinda et al., 2016). Ein aktuelles Beispiel stellen Gesichtserkennungssysteme dar, die Nutzen, aber auch Risiken mit sich bringen können (ISB 2023).



Abbildung 6: Die Jugendlichen bei der Arbeit mit Unravel

Fazit und Ausblick

Obwohl in Bayern bisher noch keine Klasse den neuen G9-Lehrplan der elften Jahrgangsstufe, in der die Künstliche Intelligenz verankert ist, durchlaufen hat, konnte das Konzept mit mehreren elften Klassen des alten Lehrplans in Form eines Projekts erprobt werden (Bild 6). Die Ergebnisse waren durchweg positiv. Eine abschließende Evaluation zeigte den Erfolg der Sequenz im Hinblick auf Motivation und fachlichem Lernerfolg. Gerade die intrinsische Motivation des explorativen Ansatzes fördert ein tieferes Verständnis der Funktion von Bildererkennungssystemen und Verfahren des maschinellen Lernens und ermöglicht damit eine Entmystifizierung der Künstlichen Intelligenz. Dass dieses Ziel erreicht wurde, zeigten Aussagen in den Feedbackrunden, wie „ein solches System ist eigentlich gar nicht so kompliziert“ und „da steckt ja gar nicht so viel dahinter“. Positiv war auch, dass sich selbst leistungsschwächere Schülerinnen und Schüler immer einbrachten.

Bei der geplanten Weiterentwicklung liegt der Fokus auf der Implementierung einer Exportschnittstelle für Trainingsdaten in eine CSV-Datei. Dadurch wird es möglich sein, diese Daten mit anderen Tools wie dem *Demonstrator für maschinelles Lernen* (ISB 2023) oder *Orange*⁴ zu verwenden. Somit können auch andere Verfahren des maschinellen Lernens, insbesondere k-nächste-Nachbarn- und Entscheidungsbaumalgorithmus, verwendet und deren Ergebnisse bewertet werden.

Quellen

Alle Webseiten/Links wurden zuletzt geprüft am 11.05.2023.

Brinda, T., Diethelm, I., Gemulla, R., Romeike, R., Schöning, J. und Schulte, C. (2016) Dagstuhlerklärung: Bildung in der digitalen vernetzten Welt. <https://dagstuhl.gi.de/dagstuhl-erklaerung>.

Burger W. und Burge M. (2015): Digitale Bildverarbeitung: Eine algorithmische Einführung mit Java. 3. Auflage. Springer Vieweg, Heidelberg.

Hastie, T., Tibshirani, R. und Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2. Auflage, Springer, Berlin.

Niemann, H. (1982): Pattern Analysis, Springer Series in Information Sciences 4. Springer, Berlin.

Seegerer, S., Lindner, A. und Romeike, R. (2019): AI Unplugged – Wir ziehen Künstlicher Intelligenz den Stecker. In: Pasternak, A. (Hrsg.), Informatik für alle. Bonn: Gesellschaft für Informatik.

Staatsinstitut für Schulqualität und Bildungsforschung (ISB) (2023): Der Lernbereich „Künstliche Intelligenz“ in der Jahrgangsstufe 11 des Gymnasiums, Bayerisches Staatsministerium für Unterricht und Kultus Abteilung Gymnasium.

⁴ <https://orangedatamining.com/>

Lizenz



Dieser Artikel steht unter der Lizenz CC BY 4.0 zur Verfügung.

Kontakt

Christoph Gräßl

Donau-Gymnasium Kelheim

E-Mail: c.graessl@donau-gymnasium.de