



Hochschule Neubrandenburg
University of Applied Sciences

Fachbereich Agrarwirtschaft und Lebensmittelwissenschaften
Studiengang Agrarwirtschaft

Bachelorthesis

zur Erlangung des akademischen Grades

"Bachelor of Science (B.Sc.)"

**Training und Evaluierung des neuronalen Netzes YOLOv8 zur
Detektion der Körperhaltung von Milchrindern in der Gruppe.**

vorgelegt von:

Mathilda Völkel

Erstprüferin: Prof. Dr. Lisa Bachmann

Zweitprüferin: M.Sc. Sarah Jahn

Neubrandenburg

24.04.2026

urn:nbn:de:gbv:519-thesis-2026-0263-1

Inhaltsverzeichnis

Inhaltsverzeichnis.....	I
Abbildungsverzeichnis.....	III
Tabellenverzeichnis.....	IV
Formelverzeichnis	V
Abkürzungsverzeichnis	V
1 Einleitung	6
2 Stand des Wissens	8
2.1 Tierüberwachung	8
2.2 Körperhaltungen in der Tierüberwachung.....	9
2.3 Automatisierte Tierüberwachung	11
2.4 Deep-Learning-Netzwerke in der Tierüberwachung.....	12
2.5 Objektdetektion mit neuronalen Netzen	14
3 Material.....	19
3.1 Haltungs- und Umweltbedingungen	19
3.2 Aufnahmen.....	19
3.3 Hardware	20
3.4 Software.....	20
3.4.1 YOLOv8.....	20
3.4.2 Klassifizierungs-Software Label Studio.....	21

3.4.3	Trainings-Software Jupyter Notebook.....	21
4	Methoden.....	22
4.1	Screenshots	22
4.2	Klassifizierung.....	23
4.3	Durchführung des Trainings.....	26
4.4	Validierung des trainierten Modells	28
5	Ergebnisse.....	29
5.1	Ergebnisse der Selbstüberprüfung.....	29
5.2	Ergebnisse des Trainings	30
5.3	Ergebnisse der Validierung	35
6	Diskussion	36
6.1	Bewertung der Selbstüberprüfung	36
6.2	Bewertung des Trainings	38
6.3	Bewertung der Validierungsergebnisse	41
6.4	Gesamtbewertung.....	47
6.5	Limitationen der Arbeit	48
6.6	Schlussfolgerung	49
6.7	Ausblick.....	49
7	Literaturverzeichnis.....	51
8	Anhang	58
9	Eidesstattliche Erklärung	72

Abbildungsverzeichnis

Abbildung 1: Beispiele für verschiedene Sensoren, die am Rind und im Stall genutzt werden (Symbolbild, generiert mit ChatGPT).	11
Abbildung 2: Schematischer Aufbau von CNNs wie den YOLO-Modellen, verändert nach Altenberger & Lenz [33].	14
Abbildung 3: Struktur der YOLO-Netzwerke, verändert nach J. Redmon [35].	15
Abbildung 4: Skizze des Versuchsaufbaus auf dem Milchviehbetrieb.	19
Abbildung 5: Labeling-Oberfläche des Programmes Label-Studio.	23
Abbildung 6: Die typischen Körperhaltungen in den vier Klassen.	25
Abbildung 7: Verteilung der Labels auf die vier Klassen im Trainingsdatensatz.	30
Abbildung 8: Verlauf der Loss-Kurven in Training 8.	32
Abbildung 9: Die Entwicklung der mAP im Trainingsverlauf von 100 Epochen.	32
Abbildung 10: Precision-Recall-Kurve aus Training 8.	33
Abbildung 11: Fehlerbeispiel Labeln: eine Kuh am Rand des Bildes ist von YOLOv8 korrekt detektiert, aber nicht als „Stehend“ klassifiziert (grün), sondern als „Liegend seitlich“ (gelb).	37
Abbildung 12: Fehlerbeispiel Validierung: falsch negative Detektionen (oben) und falsch positive Detektionen (unten) in den Klassen "Liegend eingerollt" (links) und "Liegend seitlich" (rechts).	42
Abbildung 13: Fehlerbeispiel Validierung: eng stehende Kühe (links) und Kameraperspektive von oben (rechts).	43

Tabellenverzeichnis

Tabelle 1: Anzahl der Screenshots pro Kamera.....	22
Tabelle 2: Definitionen und Abgrenzungen der vier Klassen.....	24
Tabelle 3: Einstellungen der Trainingsdurchläufe.	27
Tabelle 4: Ergebnisse der Validierung des Trainingsdatensatzes.....	29
Tabelle 5: Ergebnisse der 8 Trainingsdurchläufe, AP für jede Klasse und mAP für alle Klassen in Prozent.	31
Tabelle 6: Übersicht der Leistungskennzahlen aus dem Training. Precision und Recall für jede Klasse und der Mittelwert für alle Klassen.	34
Tabelle 7: Übersicht der Leistungskennzahlen aus der Validierung. Precision und Recall für jede Klasse und der Mittelwert für alle Klassen.	35

Formelverzeichnis

Formel 1: Intersection over Union, verändert nach Rezatofighi et al. [38]	16
Formel 2: Berechnung der Precision nach Gong et al. [40]	17
Formel 3: Berechnung des Recalls nach Gong et al. [40]	17
Formel 4: Berechnung der Übereinstimmung in der Selbstüberprüfung.....	26

Abkürzungsverzeichnis

AP	Average Precision (mittlere Detektionsleistung)
BCS	Body Condition Score
CNN	Convolutional Neural Network
FN	False negative (falsch Negativ)
FP	False positive (falsch Positiv)
Fps	Frames per Second (Bilder pro Sekunde)
IoU	Intersection over Union (Schnittmenge über Vereinigung)
mAP	Mean Average Precision (mittlere Detektionsleistung über alle Klassen)
PLF	Precision Livestock Farming
TN	True negative (richtig Negativ)
TP	True positive (richtig Positiv)

1 Einleitung

Die Tierüberwachung wird in der modernen Tierhaltung immer wichtiger. Sie wird herkömmlicherweise vom Tierhalter oder geschulten Mitarbeitern durchgeführt. Dadurch ist sie zweitaufwendig, stark vom Personal abhängig und subjektiv [1]. Der Ansatz des Precision Livestock Farming hingegen strebt eine dauerhafte Echtzeitüberwachung der Nutztiere mit technischen Hilfsmitteln an. So werden Gesundheit, Wohlbefinden, Leistungsparameter und Fortpflanzung objektiv und präzise überwacht [2].

Es gibt viele Vorschläge für Parameter, die im Precision Livestock Farming überwacht werden können [3]. Ein Parameter, der Aufschluss über die Gesundheit und das Wohlbefinden bei Rindern geben kann, ist die Körperhaltung. Das Liege- und Stehverhalten sowie die Liegeposition geben Hinweise auf Krankheitsgeschehen, Entspannung oder Stress und die Bequemlichkeit der Stallumgebung [4, 5, 6].

Innerhalb des Precision Livestock Farming entwickelten sich Ansätze zur Tierüberwachung mit künstlicher Intelligenz, sogenannten Deep-Learning-Netzwerken [7, 8]. Diese werten Bilddaten aus, die zuvor mit einer Kamera im Stall erstellt wurden. Diese Technologie ist attraktiv, weil sie im Stall nicht-invasiv ist und Objektivität und Präzision verspricht [2]. Einige Deep-Learning-Netzwerke sind auch zur Echtzeitüberwachung geeignet. Dazu zählen die von der Ultralytics Inc. entwickelten YOLO-Modelle [9].

Zwei vorangegangene Arbeiten haben gezeigt, dass YOLO-Modelle die Körperhaltung von Rindern erkennen können. Diese Forschungen beschäftigten sich mit der Klassifizierung von ein oder zwei Tieren und eindeutigen Verhaltenskategorien, z.B. Stehen oder Liegen. [10, 11]. Ob die Detektion in Gruppen mit mehreren Tieren und Körperhaltungen ebenfalls gut funktioniert, ist bisher nicht getestet worden. Auch die Detektionsleistung bei Körperhaltungen, die schwerer voneinander zu unterscheiden sind als bspw. die Kategorien „stehend“ und „liegend“, ist bisher nicht getestet worden. Daher ist das Ziel der vorliegenden Arbeit, das neuronale Netz YOLOv8 zur Detektion der Körperhaltung von Milchrindern in der Gruppe zu trainieren um zu überprüfen, ob bei der Detektion in der Gruppe ähnlich gute Leistungen erreicht werden, wie in

Kleingruppen und ob die geringe Unterscheidbarkeit der Körperhaltungen eine Rolle spielt. Dieser Versuch soll Hinweise auf die Umsetzbarkeit, Genauigkeit und Grenzen der Detektion in Gruppen geben.

2 Stand des Wissens

2.1 Tierüberwachung

Tierüberwachung ist in allen landwirtschaftlichen Produktionsrichtungen essentiell. Wohlbefinden, Gesundheit und Leistungssteigerung gehen Hand in Hand und können nur durch Tierüberwachung gesichert werden [12]. Werden Stress und erste Krankheitsanzeichen wie Veränderungen in den Bewegungsabläufen oder im Fressverhalten erkannt, ergibt sich die Chance, einzugreifen und schwere Krankheitsverläufe zu verhindern [3, 13].

Die meisten Ansätze zur Tierüberwachung basieren auf manuellen, regelmäßigen Inspektionen und Bewertungen [3]. Die Beurteilung des Tierwohls anhand von Verhaltensparametern obliegt dem Menschen und wird häufig manuell bzw. visuell durchgeführt [14]. Dazu ist ein tiefgreifendes Verständnis des tierischen Verhaltens notwendig, die Methode ist zudem subjektiv, zeitaufwendig und lässt Interpretationsspielraum [2].

Tierwohl beschreibt den Zustand eines Tieres im Hinblick auf sein Wohlergehen und seine Bedürfnisse und wird betrachtet unter den Aspekten der Gesundheit, der Ausführung natürlicher Verhaltensweisen und dem emotionalen Zustand. [15]. Das Tierwohl lässt sich mit verschiedenen Parametern messen. Das Nationale Tierwohlmonitoring schlägt neben haltungsbezogenen Parametern wie Wasserversorgung oder Tier-Fressplatz-Verhältnis auch tierbezogene Parameter vor, unter anderem Körperkondition, Verschmutzung oder Schwanzschäden. Daneben werden Daten aus Management-Programmen sowie Befunde vom Schlachthof herangezogen [14]. Unter den tierbezogenen Faktoren befinden sich Lahmheit und das Aufstehverhalten, unter den haltungsbezogenen Faktoren die Beschaffenheit der Liegefläche und das Vorhandensein von Einstreu [12]. Letzterer Faktor ist Tierwohl-relevant, weil lange Ruhephasen sich positiv auf die Tiergesundheit auswirken. Nicht-flüssiges Aufstehverhalten deutet auf die Überforderung der Anpassungsfähigkeit der Rinder hin, die Unsicherheit, Stress und Schmerzen zur Folge haben kann [14].

Ist also das Ziel, eine möglichst angenehme Umgebung für gehaltene Tiere zu schaffen, müssen Bewegungsabläufe und Körperhaltungen ermittelt und ausgewertet werden, weil durch sie direkt auf das Wohlbefinden der Tiere in einer Haltungsumgebung geschlossen werden kann [5, 6].

2.2 Körperhaltungen in der Tierüberwachung

Nach van Erp-van der Kooij et al. (2019) sind Steh- und Liegeverhalten und verschiedene Liegepositionen multifaktoriell beeinflusste Indikatoren für Wohlbefinden und Gesundheit eines Rindes oder einer Herde [16].

Die häufigste Liegeposition ist die Bauchlage. Die Beine können in verschiedene Richtungen ausgestreckt werden. Der Kopf wird meist aufrecht gehalten. Er kann auch nach vorn oder zur Seite an die Rippen abgelegt werden. In der Seitenlage lagert das Gewicht der Kuh auf Flanke und Schulter, der Kopf kann angehoben oder abgelegt werden. Jüngere Kühe bevorzugen diese Position [17].

Eine ausreichende Liegezeit wird häufig als wichtiger Aspekt der Sicherung von Wohlbefinden und Leistung bei Milchrindern angesehen [5]. Kuhindividuelle Faktoren wie Alter, Milchleistung und Reproduktionsstatus beeinflussen diese ebenso wie der Tagesablauf mit Fütterungs- und Melkzeiten [18]. Lahmheit führt tendenziell zu verlängerten Liegezeiten [6], während Mastitis sie eher verkürzt [13]. In Haltungssystemen mit Freiflächen, also in der Weidehaltung und auf Tiefstreu-Freiliegeflächen liegen Rinder durchschnittlich 9 Stunden am Tag, während sie in Anbindehaltung und Liegeboxen-Laufställen im Schnitt 10-12 Stunden liegen [5]. In solchen Haltungsformen verkürzen jedoch ungünstige Konditionen wie zu wenig oder unbequeme und kleine Liegeplätze [19] oder Hitze und Regen die Liegezeit. Der Effekt verlängerter Liegezeiten älterer Kühe oder verkürzter Liegezeiten in Weidehaltung auf das Tierwohl ist noch nicht bekannt. Deswegen muss das Liegeverhalten immer vor dem Hintergrund der betriebsindividuellen Gegebenheiten beurteilt werden [5].

Die Liegeposition wird von verschiedenen Tierwohl-Parametern beeinflusst und lässt Rückschlüsse auf sie zu. In einer Studie an drei niederländischen Universitäten [16] wurden die Körperhaltungen von liegenden Rindern auf Weideflächen und in Liegeboxen verglichen. Es wurden lange und breite Liegepositionen unterschieden. In

breiten Positionen werden Gliedmaßen zur Seite ausgestreckt, in langen Positionen nach vorne oder hinten. In schmalen Positionen sind die Gliedmaßen nah am Körper. Die Rinder auf Weide- oder Freiliegeflächen nahmen vielfältigere und eher lang gestreckte oder breite Positionen ein, bei denen Hals oder Beine weit ausgestreckt wurden. In Liegeboxen wurden eher schmale, in Boxen mit weichem Boden eher lange Positionen beobachtet. Breiten Positionen wurden in Boxen mit Trennbügeln beobachtet. Je weicher der Untergrund einer Freiliegefläche, desto höher waren die Liegezeiten und flexibler die Positionen. Feuchter Untergrund wurde eher gemieden. Über die Liegeposition in Boxen oder auf anderem Untergrund lassen sich also Rückschlüsse auf ihre Eignung für eine Herde ziehen [16].

Stehverhalten ist bei Rindern mit Fortbewegung, der Aufnahme von Futter und Wasser sowie anderen Verhaltensweisen wie Sozialverhalten verknüpft [17]. Das Liege-, Steh- und Aufstehverhalten kann bei tragenden Kühen einen deutlichen Hinweis auf eine bevorstehende Kalbung liefern. Die Anzahl der Haltungswechsel von liegend zu stehend und umgekehrt nimmt mit Nähe zur Geburt deutlich zu [20]. Anhand der Häufigkeit einer Liegeposition mit ausgestrecktem Hinterbein lässt sich die Wahrscheinlichkeit einer Geburtskomplikation vorhersagen [20].

Des Weiteren beeinflusst das Stressgeschehen das Liege- und Stehverhalten in einer Herde. Generell neigen Rinder dazu, bei Hitzestress aufzustehen. Auch die bevorzugten Liegepositionen verändern sich, noch bevor die Rinder mit Aufstehen reagieren. Die Kühe liegen deutlich seltener mit dem Kopf seitlich abgelegt, sie halten ihn eher gerade und hoch [4].

2.3 Automatisierte Tierüberwachung

Precision Livestock Farming (PLF) ist ein Ansatz in der Tierhaltung, der die dauerhafte Echtzeitüberwachung von Gesundheit, Wohlbefinden, Produktion und Fortpflanzung zum Ziel hat [1].

Bei Rindern ist die automatische Überwachung mit technischen Hilfsmitteln weit verbreitet und wird genutzt, um den Reproduktionszyklus und die Gesundheit zu überwachen. Zu solchen Technologien zählen unter anderem Pedometer sowie Halsbänder und Ohrensensoren, GPS-Überwachung, mit Sensoren ausgestattete Boli und alle Sensoren, die in Melktechnik verbaut sind [3]. Abbildung 1 zeigt beispielhaft die Position einiger Sensoren an der Kuh.



Abbildung 1: Beispiele für verschiedene Sensoren, die am Rind und im Stall genutzt werden (Symbolbild, generiert mit ChatGPT).

Aktivitäten des Rindes können über Accelerometer erfasst werden. Diese 3D-Beschleunigungssensoren erfassen die Bewegungen der Kuh. Aus diesen Bewegungsdaten wird auf ein Verhalten geschlossen, bpsw. laufen oder liegen [21]. Pedometer werden in der Brunsterkennung genutzt. Sie sind an den Beinen der Rinder befestigt und erfassen die Schrittzahl [22]. Sensor-Halsbänder können ebenfalls die Aktivität bzw. das Verhalten erkennen. Dazu zählen beispielsweise Kauen, Wasseraufnahme und Wiederkauen [23]. Pansenboli können den Pansen-pH-Wert

und die Temperatur messen und haben Aktivitätssensoren eingebaut [24]. All diese Daten werden dem Tierbetreuer zur Verfügung gestellt, um Entscheidungen auf tierindividueller Basis treffen zu können [25].

Automatisierte Tierüberwachungssysteme sind selten extern validiert, sodass die tatsächliche Zuverlässigkeit unklar ist [3]. Bei Liege- und Stehverhalten sowie Wiederkauaktivität sind Genauigkeit der Messungen und die Validierungsrate hoch. Für aktive Bewegung, Gesundheitsindikatoren, BCS oder Mastitis-Erkennung ist die Validierungsrate geringer [3]. Die Genauigkeit der Sensoren hängt auch mit den Haltungs- und Umweltbedingungen zusammen, weil eine Technologie, die für ein Produktionssystem validiert wurde in einem anderen nicht zwangsläufig dieselben Ergebnisse liefert [26].

Darüber hinaus decken diese Systeme einige wichtige Gesundheits- und Tierwohlindikatoren nicht ab. Tierhygiene, also bspw. die Sauberkeit von Körper und Euter, soziale Interaktionen oder emotionale Zustände können mit den beschriebenen Systemen nicht erfasst werden [3].

2.4 Deep-Learning-Netzwerke in der Tierüberwachung

Deep-Learning Netzwerke können im Rahmen des PFL genutzt werden. Tieridentifikation, Verhaltensanalyse und Gesundheitsüberwachung können automatisiert werden. Einige Forschungsarbeiten haben sich bereits mit der Entwicklung solcher Netzwerke beschäftigt [10, 11, 6, 20, 27]. Solche bildverarbeitenden Algorithmen werden Computer-Vision-Modelle genannt. Ihr Ziel ist es, Bilddaten wie ein Mensch oder noch genauer auszuwerten [28].

In einer Britischen Forschungsarbeit wurde eine vollautomatische Echtzeit-Lahmheitserkennung für Kühe entwickelt, die sehr sicher funktionierte. Die Genauigkeit der Lahmheitserkennung betrug bis zu 100% und die Genauigkeit der Klassifikation des Schweregrades 94%. Die Lahmheitserkennung funktionierte für mehrere Tiere gleichzeitig, konnte also für die Detektion in Gruppen eingesetzt werden [27].

Zur Geburtsüberwachung bei Rindern wurde ein Computer-Vision-Modell entwickelt, das den Kalbezeitpunkt und den Kalbeverlauf anhand von Körperhaltungen vorhersagt. Das Modell konnte Körperteile voneinander abgrenzen und Haltungen unterscheiden. Nach dem Training lag die Durchschnittsgenauigkeit bei 84% [20].

Weiterhin war von Interesse, ob sich im Verhalten von Kühen mit leichtem und schwerem Kalbeverlauf Unterschiede zeigen. Kühe mit schweren Kalbeverläufen neigen dazu, sich schon drei Stunden vor der Kalbung mit ausgestrecktem Bein hinzulegen. Kühe mit unproblematischen Kalbungen zeigten dieses Verhalten erst ein bis zwei Stunden vorher. Mit dem Wissen um diese Auffälligkeit kann frühzeitiger in schwere Geburtsvorgänge eingegriffen werden, sodass weniger folgenreiche Komplikationen auftreten [20].

An manchen Stellen werden künstliche Intelligenz bzw. neuronale Netze bereits auf landwirtschaftlichen Betrieben eingesetzt und kommerziell vermarktet. Die britische Firma „CattleEye“ bietet ein System zur Selektion für die Klauenpflege und die Erkennung von Lahmheit [29]. Die niederländische Firma Nedap bietet mit „SmartSight“ ebenfalls ein neuronales Netz zur Lahmheitserkennung an [30]. „Livestock 3D“-Systeme aus der Schweiz schätzen darüber hinaus den BCS und das Gewicht [31]. Von diesen drei kommerziellen Modellen ist nur „CattleEye“ extern validiert worden [32].

Die Grenze dieser Systeme ist bisher die Variabilität der Umweltbedingungen [7]. Sich verändernde Lichtverhältnisse und Witterung, Schatten und Staub können die Bildqualität beeinflussen. Die Tieridentifikation in Gruppen wird erschwert, wenn sich Tiere verdecken oder sich optisch sehr ähnlich sehen. Außerdem funktionieren bilderkennende neuronale Netzwerke schlechter in Umgebungen, die nicht der Trainingsumgebung entsprechen. Ist bspw. der Hintergrund oder die Rassen und Haltungsform anders als im Training, muss das Modell nachtrainiert oder feinabgestimmt werden [7].

2.5 Objektdetektion mit neuronalen Netzen

Aufbau neuronaler Netze

Ein neuronales Netz ist eine Art von künstlicher Intelligenz, die die Funktionsweise des biologischen Gehirns nachzuahmen versucht. Ein neuronales Netz besteht aus vielen Verarbeitungselementen, deren Funktion menschlichen Neuronen ähnelt. Zwischen den Verarbeitungselementen werden Verbindungen hergestellt. So entsteht die Bezeichnung „Neuronales Netz“ [33].

Durch die Verbindung zwischen zwei Neuronen werden Signale übertragen. Jede Neuronenverbindung hat außerdem ein Gewicht, das die Stärke des Signals bestimmt. Die Organisation der Neuronen und die Stärke der Verbindungen bestimmen das Endergebnis. Neuronale Netze sind so dazu in der Lage, Muster zu erkennen [34].

Es gibt verschiedene Typen neuronaler Netzwerke. YOLOv8, das für diese Arbeit genutzt wurde, kann bei den sogenannten Convolutional Neural Networks (CNNs) eingeordnet werden. Der Aufbau eines CNN ist beispielhaft in Abbildung 2 dargestellt.

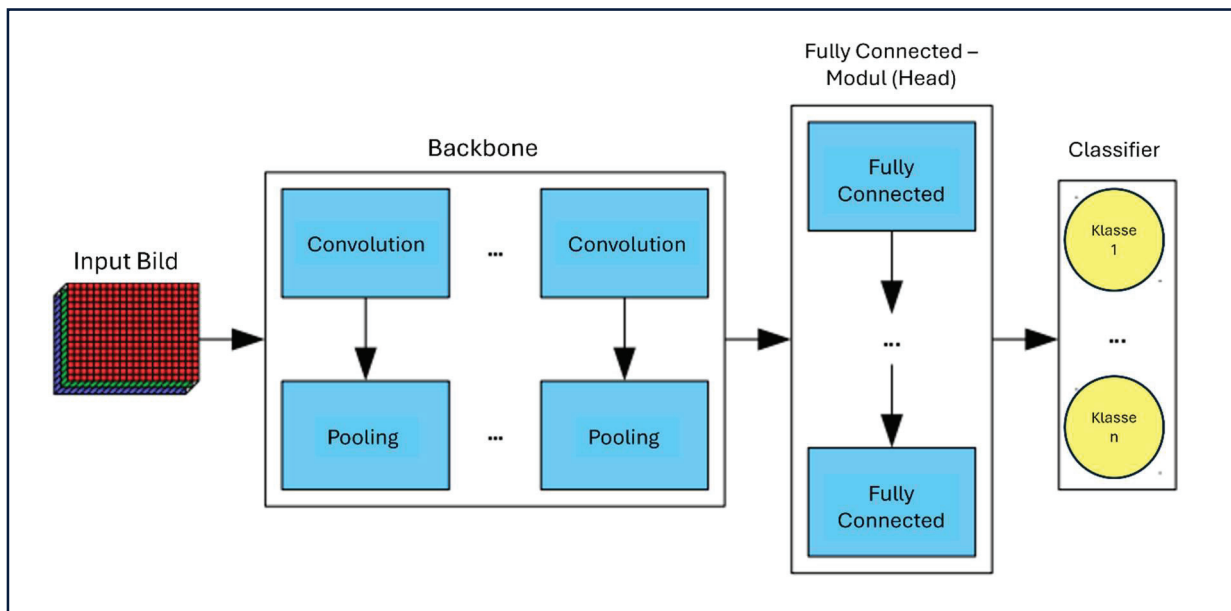


Abbildung 2: Schematischer Aufbau von CNNs wie den YOLO-Modellen, verändert nach Altenberger & Lenz [33].

CNNs unterscheiden sich von anderen Typen Neuronaler Netze (z.B. Radial Basis Function Netzwerke, Kohonen Self Organising Netzwerke) dadurch, dass die Informationen nur in eine Richtung fließen, nicht in beide [8]. Die Neuronen sind unterteilt in Inputneuronen, die Informationen aufnehmen können und Outputneuronen, die zeigen, wie das Netzwerk die Eingabeinformationen verarbeitet bzw. klassifiziert hat. Dazwischen liegen Schichten „versteckter Neuronen“, die die Informationen verarbeitet. Diese Neuronen bilden den Hauptteil des neuronalen Netzwerkes [34, 8].

Dieser zwischenliegende Teil besteht bei YOLO-Netzwerken aus „convolutional layers“ und „fully connected layers“ und ist schematisch in Abbildung 3 dargestellt. In den „convolutional layers“ fährt ein Filter aus 3x3 Pixeln das Bild ab, verstärkt alle Strukturen und extrahiert Merkmale [35]. Die entstehenden Merkmalskarten werden „gepoolt“, also gebündelt für die bessere weitere Verarbeitung [33]. In den „fully connected layers“ sind alle Neuronen miteinander verbunden. In den YOLO-Netzwerken sagen sie Wahrscheinlichkeiten und Koordinaten vorher [35].

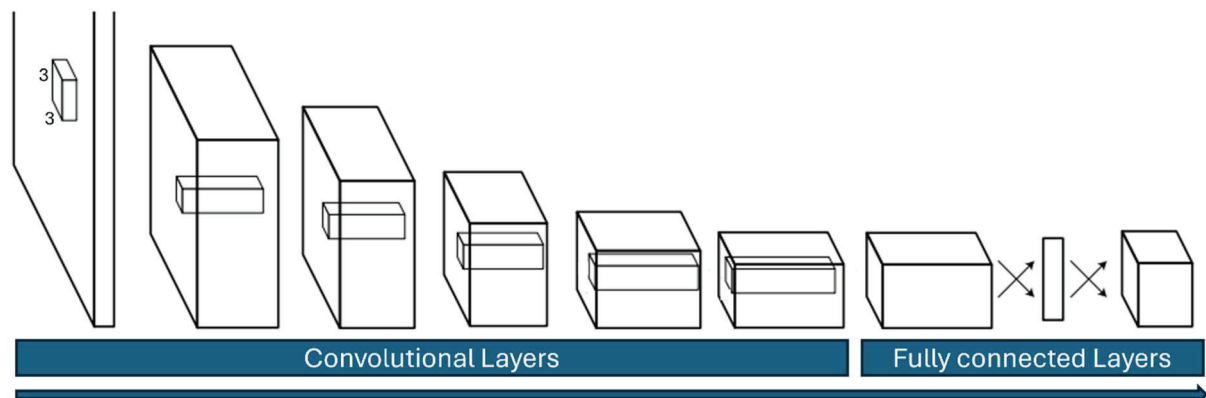


Abbildung 3: Struktur der YOLO-Netzwerke, verändert nach J. Redmon [35].

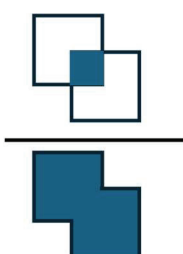
YOLOv8 ist ein verbessertes YOLO-Modell, dass 2023 herausgegeben wurde und wie alle YOLO-Modelle speziell auf die Detektion von Objekten ausgelegt ist [9]. Seine Neuronen-Schichten sind in mehreren Ebenen organisiert, genannt Backbone, Neck und Head. Die Ebene, die zuerst durchlaufen wird ist der Backbone, hier befinden sich die „convolutional layers“. Der Backbone als Grundlage des Netzwerkes liest die Bildinformation und erzeugt „Merkmalskarten“ (Feature Maps), also Versionen des Bildes, auf denen bestimmte Elemente wie Kanten, Ecken oder Linien hervorgehoben

sind. Diese werden vom Neck in verschiedenen Ebenen und Skalen kombiniert, so können unterschiedlich große Objekte besser erkannt werden. Der Head schließlich detektiert und klassifiziert die Objekte mit einer zugehörigen Wahrscheinlichkeit [36]. Er enthält die „fully connected layers“ [9]. YOLOv8 besteht also aus mehreren Ebenen und Schichten und kann mit einem Trainingsdatensatz auf Objekte angelernet werden, die nach dem Training automatisch detektiert werden. Man kann das Programm deswegen auch als Deep-Learning-Netzwerk bezeichnen [8].

Durch diese Struktur ist YOLOv8 dazu in der Lage, Objektdetektion, Instanzsegmentierung und Echtzeitverfolgung zu vereinen. Das bedeutet, es kann mehrere Objekte erkennen und voneinander abgrenzen, beispielsweise indem jedes Objekt mit einer anderen Farbe markiert wird. Darüber hinaus können Objekte in Echtzeit erkannt werden, zum Beispiel auf Videos [9].

Training neuronaler Netze

Das Training eines neuronalen Netzes dauert eine festgelegte Anzahl von Epochen. Eine Epoche stellt innerhalb eines Trainings einen Durchlauf dar, in dem alle Beispiele aus dem Trainingsdatensatz verarbeitet werden und die Leistungsparameter auf Grundlage der berechneten Fehlerrate (Loss) aktualisiert werden. Die Batch Size beschreibt die Anzahl an Elementen, die in einer Epoche verarbeitet werden müssen [37]. Für ein Training mit beispielsweise 100 Epochen und einer Batch Size von 16 werden 100 Mal 16 zufällige Bilder aus dem Trainingsdatensatz entnommen und verarbeitet. Der IoU-Wert (Intersection over Union) gibt den Grad der Überlappung zweier Formen an, in denen sich das gesuchte Objekt befindet. Diese Formen werden Bounding-Boxen genannt. Formel 1 veranschaulicht die IoU zweier Bounding-Boxen.

$$\text{IoU} = \frac{\text{Schnittmenge}}{\text{Vereinigung}} = \frac{\text{Diagramm 1}}{\text{Diagramm 2}}$$


Formel 1: Intersection over Union, verändert nach Rezatofighi et al. [38]

Der Wert befindet sich zwischen 0 (keine Überlappung) und 1 (vollständige Überlappung). Beim Training eines neuronalen Netzwerkes wird errechnet, zu welchem Grad die „ground truth“-Box, also die menschlich erstellte Bounding-Box, und die vom neuronalen Netz geschätzte „predicted“-Box überlappen [38]. Die IoU-Schwelle bedeutet, dass Bounding-Boxen, die als richtig positiv gelten sollen, mit mindestens dem Schwellwert überlappen. Andernfalls zählt die Box als negative Detektion [39].

Nach dem Training können Precision und Recall als Leistungsparameter berechnet werden. Precision bezeichnet die Genauigkeit, das heißt den Anteil richtig positiver Objekte an allen vom neuronalen Netz erkannten Objekten. Recall bezeichnet die Wiederholbarkeit, den Anteil richtig positiver Objekte an allen existierenden Objekten [40].

$$precision = \frac{TP}{TP + FP}$$

Formel 2: Berechnung der Precision nach Gong et al. [40]

$$recall = \frac{TP}{TP + FN}$$

Formel 3: Berechnung des Recalls nach Gong et al. [40]

- Wenn eine Einordnung wirklich positiv und die Vorhersage auch positiv ist, wird der Fall richtig Positiv (von engl. *true positive*, TP) genannt.
- Wenn eine Einordnung wirklich positiv und die Vorhersage negativ ist, wird der Fall falsch Positiv (von engl. *false positive*, FP) genannt.
- Wenn eine Einordnung wirklich negativ und die Vorhersage positiv ist, wird der Fall falsch Negativ (von engl. *false negative*, FN) genannt.
- Wenn eine Einordnung wirklich negativ und die Vorhersage negativ ist, wird der Fall richtig Negativ (von engl. *true negative*, TN) genannt [41].

Im Kontext des Trainings eines neuronalen Netzes mit Bilddaten würde TN bedeuten, dass Hintergrund korrekt als Hintergrund erkannt wurde. Deshalb sind falsch negative Vorhersagen sind für den vorliegende Kontext nicht relevant. TN sei hier der Vollständigkeit halber erwähnt [41].

Die Leistungsfähigkeit des Modells kann mit der Precision-Recall-Kurve grafisch dargestellt werden. Precision und Recall sind für verschiedene Konfidenzschwellen zwischen 0 und 1 abgetragen. Eine Konfidenzschwelle ist eine Mindest-Sicherheitswahrscheinlichkeit, ab der eine Vorhersage als positiver Treffer gezählt wird. Sie enthält die Wahrscheinlichkeit, dass sich die vorhergesagte Klasse in der Box befindet und wie gut die Box um das Objekt passt [35].

Die aus dem Flächeninhalt unter der Precision-Recall-Kurve berechneten AP-Werte (Average Precision) gehören standardmäßig zu den Leistungsparametern, die an eine Objektdetektions-Software angelegt werden [39]. Die AP-Werte und der mAP-Wert als Mittelwert der AP-Werte aller Klassen können genutzt werden, um die Detektionsleistung zwischen den Klassen oder das vorliegende Modell mit anderen Modellen vergleichen zu können [39]. Darüber hinaus kann der Lernverlauf des Modells grafisch dargestellt werden, wenn man die mAP am Ende jeder Trainingsepoche abträgt [42].

Das Phänomen „Overfitting“ kann nach dem Training die Leistungsfähigkeit eines Modells beeinflussen. Grundlegend bedeutet eine „Überanpassung“ oder das „Overfitten“ eines Modells, dass die Trainingsdaten akkurat wiedergegeben werden können, aber das Modell sich nicht gut auf neue Daten desselben Datensatzes verallgemeinern lässt, weil einige in den Trainingsdaten gelernte Muster nicht auf die Datengesamtheit angewendet werden können. Mit anderen Worten ist das Modell zu leistungsfähig für die vorliegenden Daten und Anwendungen und lernt deshalb im Training Features, die nicht zum Objekt gehören, also beispielsweise Hintergründe [43].

3 Material

3.1 Haltungs- und Umweltbedingungen

Die Aufnahmen, die als Datengrundlage für die vorliegende Arbeit dienen, stammen aus der Tierüberwachung eines Milchviehbetriebes in Mecklenburg-Vorpommern. Der Betrieb melkt ca. 1.600 Milchrinder und betreibt eigenen Futterbau. Die Kameras waren in einer Tiefstreu-Box im Offenstall des Betriebes angebracht. Dort waren die frisch abgekalbten und lahnenden Tiere aller Altersklassen auf einer Freiliegefläche von ca. 56 x 10 m aufgestellt. Ein Skizze des überwachten Teiles der Tiefstreu-Box ist Abbildung 4 zu entnehmen. Gemolken wurde zweimal am Tag, um 5:30 und um 16:00. Von der Box aus waren teilweise auch benachbarte und gegenüberliegende Boxen zu sehen.

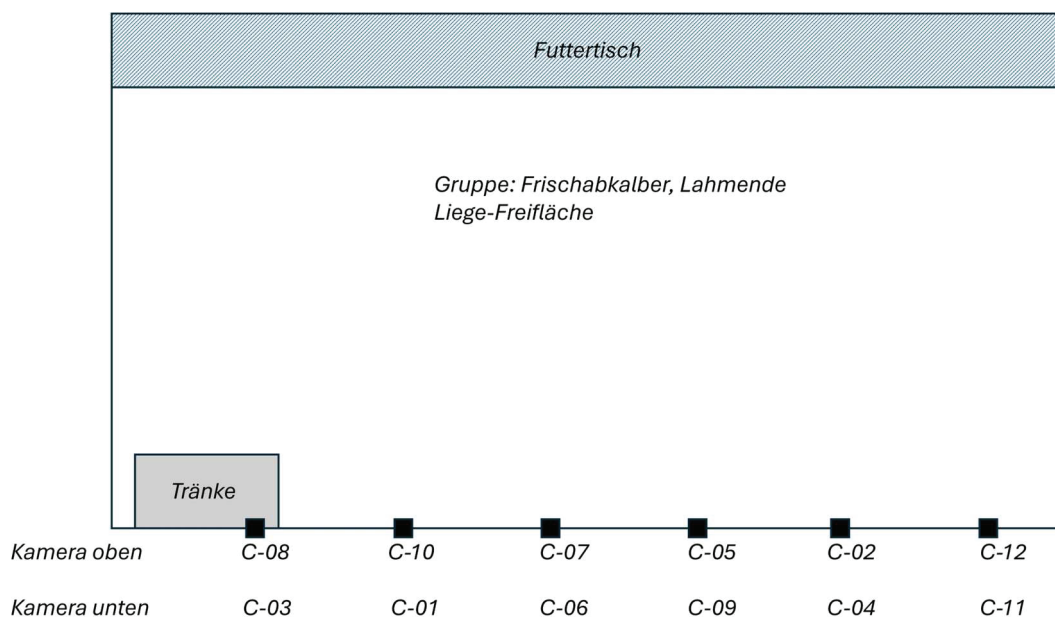


Abbildung 4: Skizze des Versuchsaufbaus auf dem Milchviehbetrieb.

3.2 Aufnahmen

Die Video-Aufnahmen, die in dieser Arbeit verwertet wurde, stammen von 12 Kameras aus dem Zeitraum vom 13.06.24 bis Tag 27.06.24. Die Kameras nahmen 24 Stunden am Tag auf. Die Roh-Videos sind jeweils 30 min lang und wurden mit 10 fps mit einer Auflösung von 1920x1080 aufgenommen. Die Kameras wurden über die halbe Länge

der Box (ca. 27 m) an 6 verschiedenen Positionen angebracht, jeweils auf 2,2 m und 3,7 m Höhe. Die genaue Position und Bezeichnung der Kameras ist in Abbildung 4 skizziert. Sie schalteten bei geringer Beleuchtung in einen Infrarot-Modus, um Nachtaufnahmen zu erzeugen. Dadurch konnten unterschiedliche Lichtverhältnisse berücksichtigt werden.

3.3 Hardware

Training und externe Validierung wurden auf einem Linux Red Hat 8.0 Betriebssystem mit AMD EPYC 74F3 24-Core Prozessor mit 3.2 GHz CPU Base-clock, 528 GB RAM sowie 4 NVIDIA-Grafikkarten A100-SXM4-80GB mit insgesamt 320 GB Video RAM (Nvidia, Santa Clara, CA, USA) ausgeführt [10].

Für die Aufnahmen im Versuchsstall wurden 12 Hikvision Dome Kameras Modell HiWatch HWI-D140H-M verwendet, die mit zwei Switches verbunden waren (8-port Gigabit Ethernet PoE + Unmanaged Switch), Modell GS108PP (NETGEAR International Limited, Cork, Ireland). Die Switches waren mit einem Mini-PC gekoppelt (Cirrus7 nimbin) der mit Linux Mint 21.2 Victoria lief. Die Videos wurden direkt auf einer externen Festplatte gespeichert (SEAGATE Extension HDD 5TB).

3.4 Software

3.4.1 YOLOv8

In dieser Arbeit wurde das von Ultralytics Inc. entwickelte neuronale Netz YOLOv8 in den Varianten YOLOv8s, -n, -m und -l genutzt (Version: YOLOv8 8.3.203, Ultralytics Inc., Frederick, MD, USA). YOLO ist eine Abkürzung und steht für „You Only Look Once“. Dieser Name spielt auf den Zweck der YOLO-Netze an, die auf Objekte trainiert zu werden, sodass die Objekterkennung danach automatisiert ablaufen kann.

3.4.2 Klassifizierungs-Software Label Studio

Die Open Source Plattform Label Studio (HumanSignal Inc., Version: 1.15.0, San Francisco, CA, USA) dient zum Labeln.

Label Studio ist in der Lage, die Koordinaten der Bounding-Boxen für jedes Bild zu speichern und als txt-Dateien auszugeben. In diesem Format können sie von YOLO ausgelesen und für das Training verwendet werden.

3.4.3 Trainings-Software Jupyter Notebook

Die Software Jupyter® Notebook ist eine webbasierte Computerplattform für verschiedenartige Rechenaufgaben. Das Entwicklungsprojekt ist quelloffen und gemeinnützig und wird gestützt von der Linux Foundation [44]. YOLOv8 wurde in der Jupyter-Notebook-Umgebung auf die zuvor gelabelten Bilder angelernet und die Trainingsergebnisse und die trainierten Modelle gespeichert.

4 Methoden

4.1 Screenshots

Aus den Videos der Kameras 1 bis 6 sowie der Kameras 9 bis 12 wurden Screenshots aufgenommen, die als Trainings-Material für Yolov8 dienen sollten. Aus prozesstechnischen Gründen wurden Kamera 7 und 8 noch nicht ausgewertet. Die Screenshots hat eine studentische Hilfskraft der Hochschule Neubrandenburg erstellt. Pro Kamera wurden zufällig einige Videos ausgewählt und unterschiedlich viele Bilder herausgeschnitten, jeweils aus Tagaufnahmen (farbig) und Nachtaufnahmen (schwarz-weiß), um unterschiedliche Belichtungsverhältnisse darzustellen. Die absolute Anzahl der Bilder pro Kamera kann Tabelle 1 entnommen werden. Aufnahmen der leeren Box zu den Melkzeiten wurden nicht verwertet. Kühe in anderen Boxen wurden nur berücksichtigt, wenn sie nicht von Fressgittern verdeckt, sondern vollständig erkennbar waren.

Tabelle 1: Anzahl der Screenshots pro Kamera.

Kamera Nr.	1	2	3	4	5	6
Anzahl Screenshots	538	41	30	52	130	118
Kamera Nr.	7	8	9	10	11	12
Anzahl Screenshots	54	49	132	86	84	36

Im Laufe der Datenaufbereitung wurde festgestellt, dass die Körperhaltungen „Liegend seitlich“ und „Liegend eingerollt“ unterdurchschnittlich vertreten waren. Um diese Tendenz auszugleichen wurden weitere 103 Screenshots aus dem Videomaterial von Kamera 7 und 8 erstellt. Damit umfasste der Trainingsdatensatz insgesamt 1350 Bilder. Die nachgesuchten Bilder stammten hauptsächlich aus den Nachtstunden, weil

dann die Positionen „Liegend seitlich“ und „Liegend eingerollt“ am häufigsten vertreten waren. Es wurden nur solche Bilder ausgewählt, in denen mindestens zwei Kühe eine der gewünschten Haltungen hatten. Sowohl farbige als auch schwarz-weiße Bilder wurden einbezogen.

4.2 Klassifizierung

Die Screenshots aller Kameras wurden im nächsten Schritt in Label Studio händisch klassifiziert. Alle erkennbaren Kühe in den Bildern wurden durch Bounding-Boxen markiert und einer der vier Klassen „Stehend“, „Liegend gerade“, „Liegend eingerollt“ und „Liegend seitlich“ zugeordnet. Jede der vier Klassen steht für eine Körperhaltung. Die Klassifizierung erfolgt, indem eine Bounding-Box möglichst eng um die gesamte Kuh gezogen und einer Klasse zugeordnet wird. Die Zuordnung zu einer Klasse wird durch farbliche Markierung angezeigt. Dieser Vorgang wird im Folgenden „Labeln“ genannt. Auf die klassifizierten Bounding-Boxen wird als „Labels“ Bezug genommen. Abbildung 5 zeigt beispielhaft, wie ein gelabeltes Bild in der Label-Studio-Oberfläche aussehen kann.



Abbildung 5: Labeling-Oberfläche des Programmes Label-Studio.

Um alle Körperhaltungen und Zwischenformen eindeutig einer der vier Klassen zuzuordnen, wurden die Tabelle 2 zu entnehmenden Definitionen und Abgrenzungen der Körperhaltungen verwendet.

Tabelle 2: Definitionen und Abgrenzungen der vier Klassen.

Stehend	Alle Kühe, die den Boden nur auf den Klauen berühren sowie alle Kühe, die sich im Ablegen oder Aufstehen befinden, ab dem ersten Kopfschwung. Die typische „stehende“ Kuh steht still oder läuft.
Liegend gerade	Alle Kühe, die in Brustlage liegen. Wamme und Brust berühren den Boden, der Kopf wird gerade nach vorn oder leicht zur Seite gehalten und ist nicht abgelegt. Grenzformen von „Liegend eingerollt“ und „Liegend seitlich“ wurden im Zweifelsfall als „Liegend gerade“ klassifiziert. Die typische als „Liegend gerade“ bewertete Kuh liegt mit den Vorderbeinen eingeklappt und der Hüfte seitlich weggedreht auf der Brust und trägt ihren Kopf hoch.
Liegend eingerollt	Alle Kühe, die in Brustlage den Kopf seitlich angeklappt oder abgelegt haben. Der Kopf muss abgelegt sein und der Winkel zur Linie der Wirbelsäule weniger als 90° betragen. Die typische „eingerollte“ Kuh liegt mit ihrem Kopf an den Rippen.
Liegend seitlich	Alle Kühe, deren Brust und Wamme beim Liegen nicht mehr den Boden berührt und deren Gewicht nicht auf einem der Vorderbeine lagert. Die typische „seitlich“ liegende Kuh liegt gerade und flach mit Schulter und Seite auf dem Boden, die vier Beine gerade von sich gestreckt.

Abbildung 6 enthält Beispiele für die klassischen Körperhaltungen jeder Klasse.



Abbildung 6: Die typischen Körperhaltungen in den vier Klassen.

Beim Klassifizieren eines Bildes wurden alle Kühe berücksichtigt, die im Bild zu sehen waren. Von anderen Kühen verdeckte Kühe oder Kühe am Rand des Bildes wurden einbezogen, sobald ihre Körperteile und Haltung klar zu erkennen waren.

Die Ergebnisse der händischen Klassifizierung sollen dem neuronalen Netz möglichst genau definieren, wie sich die einzelnen Klassen voneinander abgrenzen. Die Bounding-Boxen wurden möglichst nah um die erkannten Kühe gezogen, damit der Hintergrund die Trainingsdaten nicht aufweicht.

Um den Prozess des Labelns zu vereinfachen lief die untrainierte Basis-Version von YOLOv8 in Label Studio. Diese Version ist bereits dazu in der Lage, Objekte wie Kühe zu detektieren, weil sie mit dem COCO-Datensatz vortrainiert ist [9]. Der COCO-Datensatz („Common Objects In Context“) ist ein Trainings-Datensatz für Deep-Learning-Netzwerke und enthält verschiedene Objektkategorien. Er wird unter anderem zum Benchmarking von Computer-Vision-Modellen verwendet [45].

Um die Qualität der Trainingsdaten zu überprüfen wurde eine Selbstüberprüfung durchgeführt und die Übereinstimmung der Labels ermittelt. Nach Beendigung des Labelns und mit einigen Tagen zeitlichem Abstand wurden 15% der Bilder zufällig ausgewählt und überprüft. Nicht erkannte Kühe oder falsche Klassifizierungen wurden als „abweichende Labels“ gezählt. Daraus wurde zur Beurteilung der Übereinstimmung berechnet.

$$\text{Übereinstimmung} = \frac{\text{Abweichende Labels}}{\text{Überprüfte Labels}}$$

Formel 4: Berechnung der Übereinstimmung in der Selbstüberprüfung.

4.3 Durchführung des Trainings

Es sollte geprüft werden, ob YOLOv8 mit dem vorliegenden Bildmaterial trainiert werden kann, Körperhaltungen bei Milchrindern zu erkennen und wie gut die Objekterkennung in einer Stallumgebung mit Kühen in Gruppen funktioniert. Dazu wurde YOLOv8 mit den zuvor manuell annotierten Bildern trainiert.

Für das Training von YOLOv8 wurden die Screenshots zufällig in einen Trainings-, Validierungs- und Testdatensatz geteilt, um Datenleckage zu vermeiden, sodass die Daten für Test wirklich völlig unbekannt sind [46]. Der Validierungsdatensatz wurde in der internen Validierung verwendet, die im Training inbegriffen ist. Der Testdatensatz wurde nicht verwendet, weil eine externe Validierung durchgeführt wurde. Die Trainingsdaten mit den manuellen Labels wurden in die Trainings-Software Jupyter geladen. Für das Training wurden verschiedene Varianten von YOLOv8 ausprobiert und zusätzlich die Anzahl der Parameter Epochen und Batch Size moduliert. Die IoU-Schwelle wurde auf 0,5 eingestellt.

Die Kombination der Parameter aller Trainings sind in Tabelle 3 dargestellt. Es wurden insgesamt 8 Trainings durchgeführt, bis ein zufriedenstellendes Modell ausgewählt werden konnte. Dieses wird im Folgenden „trainiertes Modell“, „trainiertes neuronales Netz“ oder „Modell“ genannt.

Ziel war es, möglichst hohe mittlere Detektionsleistungen zu erlangen. Es wurden keine Varianten größer als YOLOv8l trainiert. Diese wären wahrscheinlich noch genauer gewesen, wurde aber als zu langsam eingeschätzt.

Tabelle 3: Einstellungen der Trainingsdurchläufe.

Training	Yolov8 Variante	Epochen	Batch Size
Train 1	Yolov8n	100	16
Train 2	Yolov8n	200	16
Train 3	Yolov8n	100	32
Train 4	Yolov8s	100	16
Train 5	Yolov8s	200	16
Train 6	Yolov8s	100	32
Train 7	Yolov8m	100	16
Train 8	Yolov8l	100	16

Zuerst wurden verschiedene Kombinationen von Epochen-Anzahl und Batch Size bei den kleineren Varianten YOLOv8n und YOLOv8s ausprobiert. Als auffiel, dass bei Epochen >100 und Batch Size >16 das Overfitting für alle Varianten stieg, wurden die größeren Modelle YOLOv8m und YOLOv8l nur noch mit 100 Epochen und Batch Size 16 trainiert. Für das beste Modell wurden Recall und Präzision berechnet. Die Konfusionsmatrix-Elemente richtig Positive (TP), falsch Positive (FP) und falsch Negative (FN) wurden dazu aus den Trainingsergebnissen ermittelt. Die Auswertung der Trainingsergebnisse erfolgte in Anlehnung an die Arbeit von Jahn et al. mit einer IoU-Schwelle von 0,5 [10].

4.4 Validierung des trainierten Modells

Zur externen Validierung der Ergebnisse des Trainings detektierte das trainierte neuronale Netz bisher unbekannte Daten aus demselben Versuchsaufbau. Aus den Aufnahmen von Kamera 4 vom 15.06.24 wurden 2,4 Stunden Videomaterial bestehend aus 72 Videos herausgeschnitten. Sie sind jeweils zwei Minuten lang, starten je Stunde immer bei 0, 20 und 40 Minuten und decken alle 24 Stunden des Tages ab.

Für jedes Video hat das trainierte neuronale Netz die Kühe auf einem Bild pro Sekunde detektiert und einer Körperhaltung zugeordnet. Die Konfidenz-Schwelle wurde auf 0,5 eingestellt. Gleichzeitig wurden die Videos manuell ausgewertet. Auch hier wurde die Anzahl der Kühe jeder Klasse für jede Sekunde ermittelt.

Im nächsten Schritt sollten die Labels des neuronalen Netzes mit der manuellen Auswertung der Videos verglichen werden. Es sollte berechnet werden, wie viele Kühe vom Menschen und dem neuronalen Netz detektiert und gleich klassifiziert worden waren.

5 Ergebnisse

5.1 Ergebnisse der Selbstüberprüfung

Im Trainingsdatensatz waren zu 2,3% händische Labels fehlerhaft, also falsch klassifiziert oder Kühe im Bild übersehen worden waren. Dass bedeutet, dass die Übereinstimmung bei 97,7% lag. Die absolute Anzahl geprüfter und abweichender Labels kann Tabelle 4 entnommen werden.

Tabelle 4: Ergebnisse der Validierung des Trainingsdatensatzes.

Summe aller geprüften Labels	2213
Summe aller abweichenden Labels	51
Falsche Labels/ Geprüfte Labels	0,02305
Angabe in %	2,3
Übereinstimmung in %	97,7

5.2 Ergebnisse des Trainings

Der gesamte manuell erstellte Trainingsdatensatz, der für alle Trainings verwendet wurde, umfasste 14.296 Labels. Die Verteilung der Labels in den Klassen kann Abbildung 7 entnommen werden.

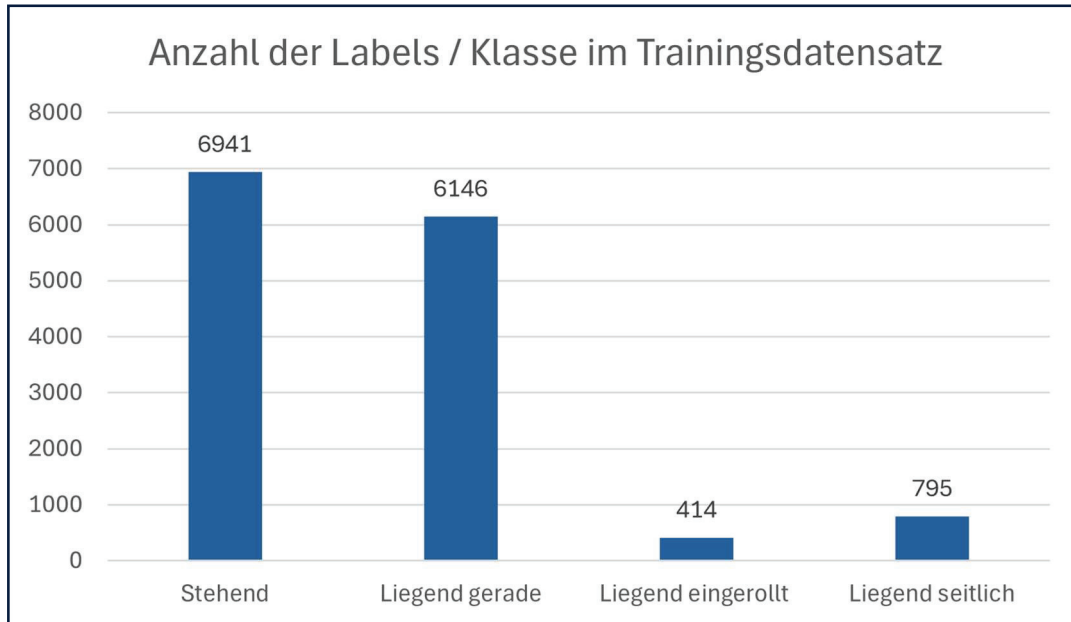


Abbildung 7: Verteilung der Labels auf die vier Klassen im Trainingsdatensatz.

In den Klassen „Stehend“ und „Liegend gerade“ waren mehr als 6.000 Labels enthalten, in den Klassen „Liegend eingerollt“ und „Liegend seitlich“ weniger als 800. In der Klasse „Stehend“ waren die meisten Labels vorhanden (6.941), in der Klasse „Liegend eingerollt“ die wenigsten (414).

Alle Ergebnisse der Trainingsdurchläufe von YOLOv8 mit der AP für jede Klasse und mAP in Prozent für alle Klassen sind in Tabelle 5 aufgeführt.

Tabelle 5: Ergebnisse der 8 Trainingsdurchläufe, AP für jede Klasse und mAP für alle Klassen in Prozent.

Training Nr.	Stehend	Liegend gerade	Liegend eingerollt	Liegend seitlich	Alle Klassen
Training 1	96	96,1	88,8	90,0	92,7
Training 2	96,9	96,9	90,7	91,0	93,9
Training 3	96,8	96,4	89,3	91,3	93,5
Training 4	96,5	96,3	89,3	91,4	93,4
Training 5	96,3	96,3	91,1	88,5	93,0
Training 6	96,3	96,1	90,5	89,5	93,1
Training 7	97,4	97,1	91,5	92,2	94,5
<i>Training 8</i>	<i>97,0</i>	<i>96,9</i>	<i>93,1</i>	<i>91,9</i>	<i>94,7</i>

Training 8 mit der Variante YOLOv8l stellte das beste Ergebnis dar, weil hier mAP und AP in den einzelnen Klassen am höchsten war. Ein zweites Kriterium war der Verlauf der Loss-Kurven. Das Modell aus Training 8 wurde ausgewählt, weil die Loss-Kurven gleichmäßig verliefen (Abbildung 8). Zum anschaulichen Vergleich ist in Anhang 1 die Loss-Kurve aus Training 5 dargestellt. Mit dem Modell, dass bei Training 8 entstand, wurde weitergearbeitet. Sein Trainingsverlauf und die Ergebnisse sollen ausführlicher dargestellt werden.

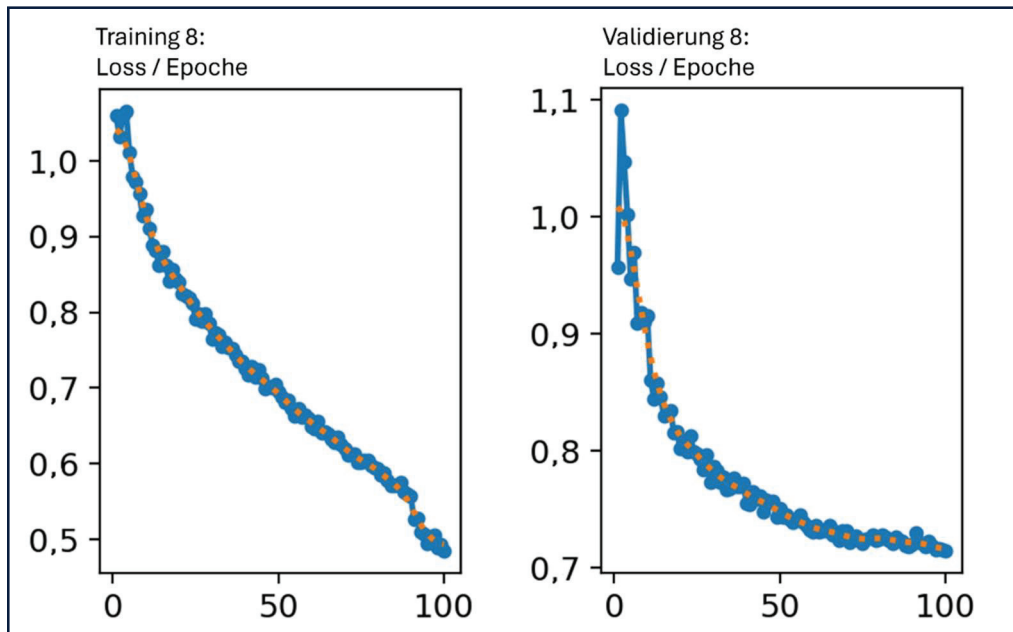


Abbildung 8: Verlauf der Loss-Kurven in Training 8.

Die Lernkurve des Modells aus Training 8 (Abbildung 9) zeigt, wie sich die mAP bei einer IoU-Schwelle von 0,5 während des Trainings entwickelte. Die Lernkurve nahm einen logarithmischen Verlauf. Die mAP stieg in den ersten Epochen schnell an, nahm dann langsamer zu und stabilisierte sich auf einem Plateau bei ca. 0,95. Dass bedeutet, dass die mAP am Ende von Epoche 100 ungefähr 95% betrug.

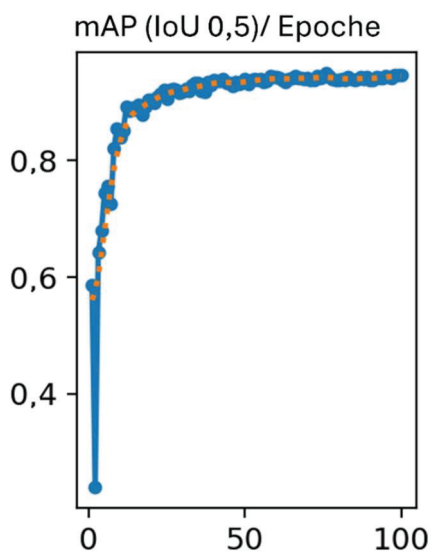


Abbildung 9: Die Entwicklung der mAP im Trainingsverlauf von 100 Epochen.

Die Precision-Recall-Kurve (Abbildung 10) gibt Aufschluss darüber, wie die Leistungsfähigkeit des Modells nach dem Training einzuschätzen ist. Aus ihr können die AP-Werte für jede Klasse und die mAP abgelesen werden. In der Klasse „Stehend“ wurde die höchste mittlere Detektionsleistung erreicht (0,970 bzw. 97%). Die Niedrigste wurde in der Klasse „Liegend seitlich“ erreicht, dort betrug sie 0,919 oder 91,9%. Die Klassen „Liegend gerade“ mit 96,6% und „Liegend eingerollt“ mit 93,1% lagen dazwischen.

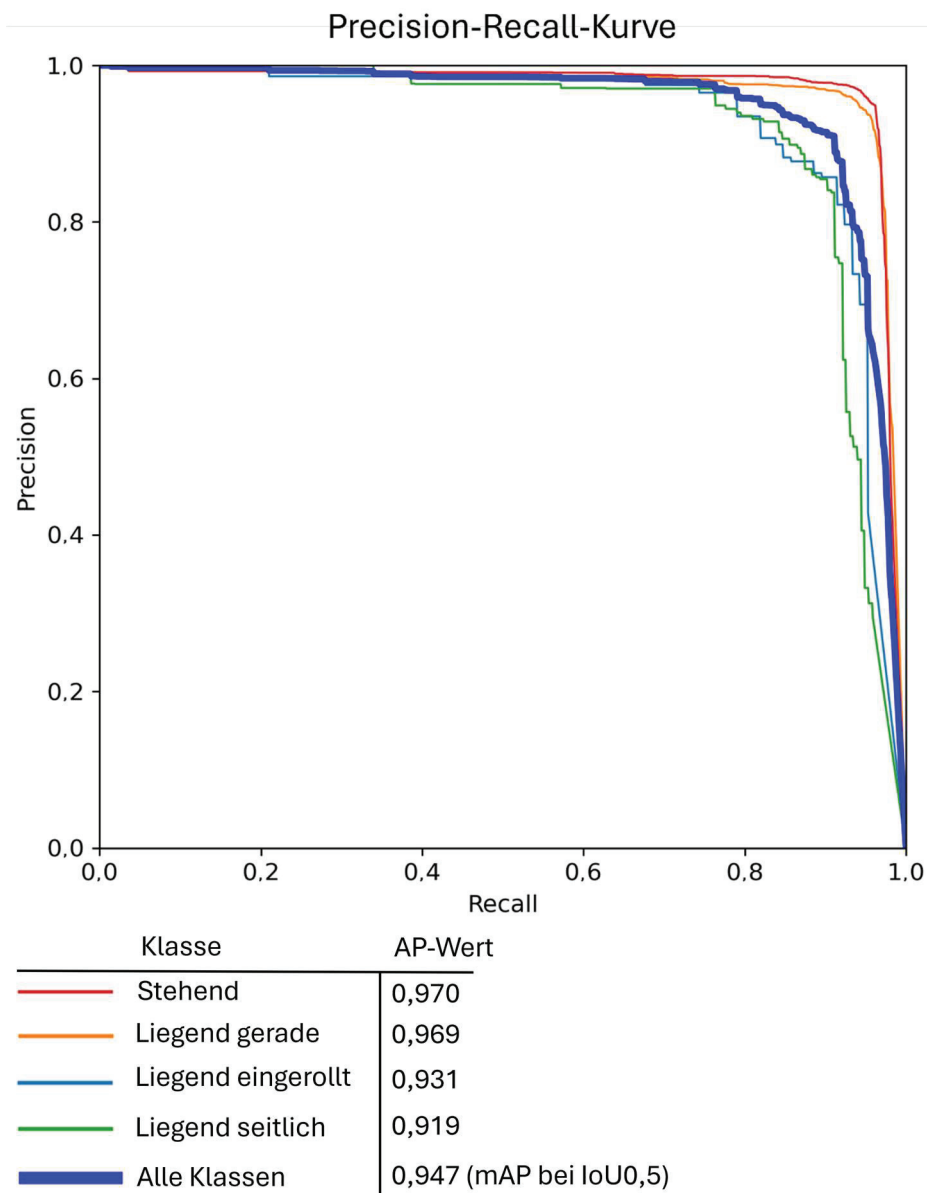


Abbildung 10: Precision-Recall-Kurve aus Training 8.

Die berechneten Werte für Präzision und Recall des Modells aus Training 8 sind für jede Klasse und für alle Klassen in Tabelle 6 dargestellt.

Tabelle 6: Übersicht der Leistungskennzahlen aus dem Training. Precision und Recall für jede Klasse und der Mittelwert für alle Klassen.

	Stehend	Liegend gerade	Liegend eingerollt	Liegend seitlich	Mittelwert alle Klassen
FP	121	150	18	47	-
FN	76	79	20	76	-
TP	1915	1669	85	187	-
Precision%	94,0	91,8	82,5	79,9	87,1
Recall%	96,2	95,5	81,0	71,1	85,9

Die Anteile der falsch positiven Vorhersagen bewegten sich zwischen 150 („Liegend gerade“) und 18 („Liegend eingerollt“). Die Spanne der falsch negativen Vorhersagen bewegte sich zwischen 79 („Liegend gerade“) und 20 („Liegend eingerollt“). Der Anteil richtig positiver Detektionen bewegte sich zwischen 1915 („Stehend“) und 85 („Liegend eingerollt“).

Die Präzision des trainierten Modells betrug 87,1%, der Recall lag bei 85,9%. Für die einzelnen Klassen lag die Präzision zwischen 79,9 und 94,0% und der Recall zwischen 71,1 und 96,2%.

5.3 Ergebnisse der Validierung

Die Ergebnisse der Validierung sind in Tabelle 7 zusammengefasst. Die Ergebnisse für alle 72 Videos können Anhang 1 entnommen werden.

Tabelle 7: Übersicht der Leistungskennzahlen aus der Validierung. Precision und Recall für jede Klasse und der Mittelwert für alle Klassen.

	Stehend	Liegend gerade	Liegend eingerollt	Liegend seitlich	Mittelwert alle Klassen
FP	36	26	9	6	-
FN	81	115	24	52	-
TP	1093	979	28	38	-
Precision%	90	87,7	38,8	44,8	65,3
Recall%	87,5	80,4	28,9	24,7	55,4

Die Ergebnisse aller 72 Videos sind hier gemittelt dargestellt. In allen Klassen wurden vom trainierten Modell mehr falsch negative Detektionen gemacht, also Kühe nicht erkannt, als falsch positive Detektionen. Die richtig positiven Detektionen hatten bei allen Klassen außer „Liegend seitlich“ die höchste Anzahl. In dieser Klasse waren 52 der Detektionen falsch negativ, also nicht erkannte Kühe. In der Klasse „Liegend eingerollt“ wurden fast so viele Kühe nicht erkannt (FN= 24) wie richtig erkannt wurden (TP=28). Die richtig positiven Detektionen lagen nur für „Stehend“ und „Liegend gerade“ bei ca. 1.000. Die Präzision für „Stehend“ lag im Schnitt bei 90%, dicht gefolgt von der Präzision für „Liegend gerade“ (87,7%). „Für Liegend eingerollt“ lag sie bei 38,8% und für „Liegend seitlich“ bei 44,8%. Der Recall lag in jeder Klassen immer leicht unter der Präzision. Für „Stehend“ ist er am höchsten (87,5%) und für „Liegend seitlich“ am niedrigsten (24,7%).

6 Diskussion

Die Trainingsergebnisse versprachen durch die mAP von 94,7% ein sehr leistungsstarkes Modell für alle Klassen. Die Precision lag zwischen 79,9 und 94,0% und der Recall zwischen 71,1 und 96,2%. In der Validierung war ein Leistungseinbruch zu beobachten. Precision und Recall sanken in allen Klassen unter die Trainingswerte. Insbesondere für „Liegend eingerollt“ und „Liegend seitlich“ sanken Precision und Recall auf unter 40%. Obwohl die Leistung im Training als hoch einzuschätzen ist, gilt die Leistung des trainierten Modelles in der Validierung als vergleichsweise moderat [7, 10, 11]. Die Ergebnisse unterliegen einem multifaktoriellen Einfluss, der im Folgenden diskutiert werden soll.

6.1 Bewertung der Selbstüberprüfung

Die Übereinstimmung in der Selbstüberprüfung lag bei 97,7%. Diese Übereinstimmung kann als hoch und zufriedenstellend eingeschätzt werden [47]. 2,3% der Labels waren falsch, entweder war eine Kuh nicht erkannt, also beim Labeln übersehen worden, oder die Körperhaltung falsch zugeordnet.

Die Qualität der Bilder war selbst bei den schwarz-weiß-Aufnahmen so gut, dass alle Kühe als solche hätten erkannt werden können. Die Klassen wurden eindeutig abgegrenzt und Zwischenformen waren selten zu finden. Die meisten Zwischenformen fanden sich im Übergang von „Liegend gerade“ zu „Liegend seitlich“. Die Handhabung von verdeckten Kühen oder abgeschnittenen Kühen am Bildrand war von Anfang an gleichbleibend. Es wurden während der Selbstüberprüfung nur zwei Label von randständigen Kühen anders bewertet. Waren viele Kühe auf einem Bild zu sehen, konnten einzelne Tiere leicht übersehen werden. Verdeckte Tiere, oder Tiere, die sich farblich weniger vom Hintergrund abhoben, wurden ebenfalls häufiger übersehen. konnten einzelne Tiere leichter übersehen werden. Verdeckte Tiere oder Tiere, die sich farblich weniger vom Hintergrund abhoben wurden ebenfalls häufiger übersehen.

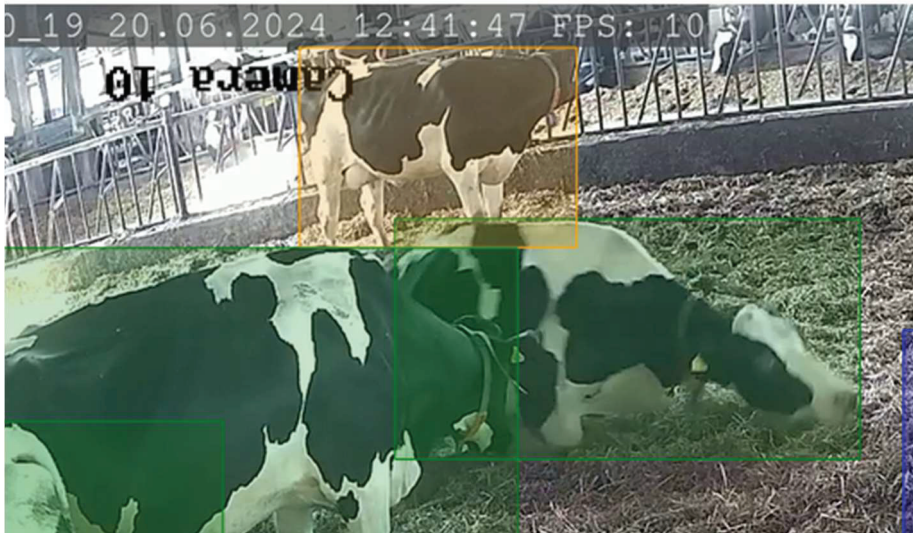


Abbildung 11: Fehlerbeispiel Labeln: eine Kuh am Rand des Bildes ist von YOLOv8 korrekt detektiert, aber nicht als „Stehend“ klassifiziert (grün), sondern als „Liegend seitlich“ (gelb).

Um das Labeln zu beschleunigen wurden alle Bilder mit YOLOv8 in der untrainierten Basis-Version vorgelabelt. Alle erkannten Kühe wurden von Label-Studio als „Liegend seitlich“ angezeigt, was vor dem Labeln festgelegt wurde. So mussten nur noch die Klassen richtig zugeordnet werden. In einigen Fällen wurden die vordetektierten „Liegend seitlich“-Labels übersehen und nicht angepasst, vor allem am Rand oder in den Ecken des Bildes. Ein Beispiel ist in Abbildung 11 dargestellt. Auch abnehmende Konzentration während des Labelns verursachte Fehler. Diese verschiedenen Fehlerquellen führten zu der beobachteten Abweichung von 2,3%, die insgesamt trotzdem als gering eingeschätzt werden kann.

Wenn falsche Labels im Training verwendet wurden bedeutet das, dass Merkmale fälschlicherweise einer anderen Klasse zugeordnet werden. Die Fehlerrate von 2,3% beeinflusst die Gesamtleistung des Modells negativ, das Ausmaß des Einflusses ist dabei mit den vorhandenen Daten nicht abzuschätzen. Fehler im Trainingsdatensatz müssen mit insgesamt mehr Trainingsdaten ausgeglichen werden. Auch eine schlechtere Generalisierung, also eine schlechtere Leistung auf neue Daten außerhalb des Trainings ist möglich [48, 49].

6.2 Bewertung des Trainings

Die vier Klassen sind im Trainingsdatensatz sehr ungleichmäßig verteilt. „Liegend eingerollt“ und „Liegend seitlich“ sind im Vergleich zu „Stehend“ und „Liegend gerade“ unterproportional vertreten. „Stehend“ klassifizierte Kühe kommen ca. achtmal häufiger vor als „Liegend eingerollt“.

Das Nachlabeln hatte den Zweck, diese im Verlauf des Labelns erkennbare Tendenz auszugleichen, war aber offensichtlich unzureichend, um die absolute Anzahl an Labels in den Klassen auf ein gleiches Niveau zu bringen. Es hat dem Trend nur etwas entgegengewirkt. In allen 103 nachgelabelten Bildern waren mindestens zwei Kühe in der Haltung „Liegend eingerollt“ oder „Liegend seitlich“ zu finden. Das bedeutet, dass von den insgesamt 1.209 Labels dieser beiden Klassen mindestens 200, also ca. 17%, durch das Nachlabeln in den Datensatz gelangten. Aus zeitlichen Gründen wurde das Nachlabeln auf ca. 100 Bilder beschränkt. Zum Ausgleich des Datensatzes mussten also noch deutlich mehr Kühe „Liegend eingerollt“ und „Liegend seitlich“ nachgelabelt werden. Diese Haltungen sind vor allem nachts zu finden, denn sie deuten auf Schlaf oder Entspannung hin [50]. Um die Varianz im Datensatz zu halten dürften allerdings nicht nur schwarz-weiß-Aufnahmen verwendet werden, sondern auch Bilder in Farbe.

Von allen Trainingsdurchläufen wurde der letzte mit der YOLO-Variante YOLOv8l für am Geeignetsten befunden. Hier wurden die höchsten AP-Werte in allen Klassen erzielt. Vor allem war ausschlaggebend, dass in den beiden weniger vertretenen Klassen „Liegend eingerollt“ und „Liegend seitlich“ die AP bei über 90% lag. Das Modell versprach deshalb einen Kompromiss aus Geschwindigkeit und Genauigkeit der Detektion in allen Klassen.

Bewertung der Lernkurve

Als erster Hinweis auf die Qualität des Modells wurde die Lernkurve herangezogen. Der rasch ansteigende Verlauf zeigt, dass das Modell die Trainingsdaten erfolgreich verarbeitet und die relevanten Features, also die Merkmale der Klassen eingrenzt und sie deshalb immer besser detektieren kann. Das Plateau bei mAP ca. 0,95 deutete darauf hin, dass das Modell bereits nach ca. 50 Epochen sein Leistungsmaximum für die Trainingsdaten erreicht hat. Danach war der Zuwachs in der mAP sehr gering.

Zuwachs bis auf ein Plateau ist ein typischer Verlauf einer Lernkurve bei neuronalen Netzen [42]. Die mAP von 0,95 am Ende der Lernkurve bei IoU 0,5 lässt sich so deuten, dass Precision und Recall hoch sind über viele Konfidenz-Schwellen, basierend auf IoU 0,5. Ein frühes Leistungsmaximum und ein langes Plateau sind ein Zeichen dafür, dass das Modell sehr gut auf die Trainingsdaten angepasst ist [42]. Es kann an dieser Stelle zu Überanpassung kommen, dann spricht man von Overfitting [43]. Für weitere Trainings könnte man erwägen, das Training verfrüht abubrechen [51].

Bewertung der Loss-Kurven

Kleinere Modelle wie YOLOv8n und YOLOv8s lieferten bei 100 Epochen und einer Batch Size von 16 eine AP von unter 90% in den Klassen „Liegend eingerollt“ und „Liegend seitlich“ und wurden deswegen verworfen. Bei Einstellung von 200 Epochen oder Batch Size 32 stieg das Overfitting der Modelle, daran erkennbar, dass der Trainings-Loss im Epochenverlauf sank, während der Validierungs-Loss stieg (vgl. Anhang 1) [51]. Das bedeutet, dass das Modell mit steigender Anpassung auf die Trainingsdaten zu spezifisch wird, also die Trainingsdaten auswendig lernt aber gelernte Muster in den Validierungs-Daten nicht wiederfindet. Deshalb werden in der Validierung die Vorhersagen ungenauer [43] und die Modelle wurden ebenfalls verworfen.

Im Gegensatz dazu verliefen die Loss-Kurven des verwendeten Modells aus Training 8 nahezu gleichmäßig fallend (vgl. Abbildung 8). Der Loss der Validierung ist kontinuierlich höher als der Trainings-Loss, steigt aber nicht an. Overfitting ist deshalb unwahrscheinlicher [43]. Es gibt allerdings keinen genauen Wert, wie groß die Schere zwischen Loss in Training und Validierung maximal sein darf. Ab einem bestimmten Abstand, abhängig von den Trainingsumständen, kann eine weite Schere zwischen Trainings- und Validierungs-Loss ebenfalls auf Overfitting hindeuten, aber auch ein Hinweis auf einen zu kleinen Trainingsdatensatz sein [51].

Bewertung der AP-Werte und der PR-Kurve

Die AP-Werte für das trainierte Modell liegen zwischen 91,9 und 97,0% und sind damit als hoch einzuschätzen. Auch die mAP von 94,7% deutet auf eine sehr gute Detektionsleistung des Modells hin, verglichen mit den AP und mAP-Werten anderer Modelle bei IoU 0,5 [7, 10, 11]. Die höchste AP wird in den Klassen „Stehend“ (97,0%) und „Liegend gerade“ (96,9%) erreicht.

Die PR-Kurve für das trainierte Modell ist im Teil Ergebnisse des Trainings in Abbildung 10 Seite 33 dargestellt. Aus der Fläche unter der PR-Kurve berechnen sich die AP-Werte von 91,9-97,0%. Der Kurvenverlauf selbst ermöglicht eine differenziertere Betrachtung der Modellleistung.

Die Precision-Werte verlaufen Plateau-ähnlich und stabil für einen großen Bereich von Recall-Werten. Das heißt, dass das Modell über einen großen Bereich an Recall-Werten eine hohe Precision beibehält, also wenig falsch positive Vorhersagen generiert. Erst bei sehr hohem Recall sinkt die Precision. Das bedeutet, dass bei hohem Recall zunehmend falsch positive Detektionen gemacht werden. In der Klasse „Liegend seitlich“ sinkt die Precision am schnellsten. Es werden also früher als in anderen Klassen mehr falsch positive Vorhersagen gemacht. Daraus kann man schließen, dass diese Klasse schwieriger zu erkennen ist oder häufiger mit anderen Klassen verwechselt wird.

Die Unterschiede bei den AP-Werten und PR-Kurvenverläufen hängen mit der Struktur des Trainingsdatensatzes zusammen. In den Klassen „Stehend“ und „Liegend gerade“, die häufig im Trainingsdatensatz vertreten waren, werden hohe Genauigkeiten erreicht. Das Modell hatte im Training genügend verschiedene Beispiele, sodass diese Klassen sicher detektiert werden können. In den weniger vertretenen Klassen „Liegend eingerollt“ und „Liegend seitlich“ sind die AP-Werte niedriger. Obwohl im Trainingsdatensatz fast doppelt so viele Kühe „Liegend seitlich“ wie „Liegend eingerollt“ gelabelt waren, ist der AP-Wert für „Liegend eingerollt“ höher (93,1%) als für „Liegend seitlich“ (91,9%). Auch diese Tatsache spricht dafür, dass seitlich liegende Kühe schwerer zu erkennen sind oder mit anderen Klassen verwechselt werden.

Grund dafür könnte sein, dass es Zwischenformen bei den Haltungen „Liegend gerade“ und „Liegend seitlich“ gibt, die für das Modell schwer zu unterscheiden sind. Ungleichmäßige subjektive Einschätzung dieser Zwischenformen während des Labelns könnte den Fehler vergrößert haben. Diese uneinheitliche Bewertung verrauscht den Trainingsdatensatz und führt zu Label-Noise, der Minderleistungen des Modells verursachen kann [49]. Des Weiteren sind die Beine seitlich liegender Kühe gut sichtbar, wodurch sie auch als stehende Kühe interpretiert werden können. Dieser Fehler ist auch in der Klassen „Liegend gerade“ bei Kühen aufgetreten, deren Beine gut sichtbar waren oder die direkt unter der Kamera lagen.

Die insgesamt sehr guten Ergebnisse bei den AP-Werten hängen auch mit der vergleichsweise niedrigen IoU-Schwelle zusammen. Für die Berechnung der AP- und mAP-Werte wurde eine IoU-Schwelle von 0,5 angesetzt. Die PR-Kurve kann genutzt werden, um für den praktischen Einsatz des Modelles eine möglichst hohe Konfidenzschwelle einzustellen, bei der Precision und Recall jedoch beide noch möglichst hoch sind.

6.3 Bewertung der Validierungsergebnisse

Vergleich von Precision und Recall in Training und Validierung

In der Validierung wurde Precision und Recall für alle Detektionen berechnet. Precision und Recall der Validierung unterscheiden sich deutlich von den Werten, die im Training berechnet wurden.

In allen vier Klassen sind Precision und Recall in der Validierung niedriger als im Training. In den Klassen „Stehend“ und „Liegend gerade“ liegen die Werte aus der Validierung für den Recall zwischen 87,5 und 80,4% und für die Precision zwischen 90 und 87,7%. Im Training dagegen lagen sie zwischen 96,2 und 95,5% für den Recall und 94 und 91,8% für die Precision. Deutlich größer ist der Unterschied in den Klassen „Liegend eingerollt“ und „Liegend seitlich“. Dort sinkt der Recall von 81 und 71,1% auf 28,9 und 24,7%. Die Precision sinkt für die beiden Klassen von 82,5 und 79,9% auf 38,8 und 44,8. Das bedeutet, dass in den Klassen mit weniger Trainingsdaten insgesamt mehr falsche Detektionen gemacht werden, von denen tendenziell mehr falsch Negative sind als falsch Positive. Falsch Negative heißt, Kühe werden „Liegend eingerollt“ und „Liegend seitlich“ detektiert, die eigentlich einer anderen Klasse

zugeordnet werden müssen. Das Modell kann die Features der Klassen „Liegend eingerollt“ und „Liegend seitlich“ nicht klar von den Features der anderen Klassen abgrenzen.

Fehlertypen des trainierten Modells

Abbildung 12 zeigt Beispiele für falsche Detektionen in den Klassen „Liegend eingerollt“ und „Liegend seitlich“. Die Features der Klasse „Liegend eingerollt“ scheint das Modell im Beispiel unten links in Abbildung 12 so zu interpretieren, dass der Kopf der Kuh vor dem Körper sein muss. Im richtigen Winkel erscheint deshalb die Kuh in Abbildung 12 als „Liegend eingerollt“.



Abbildung 12: Fehlerbeispiel Validierung: falsch negative Detektionen (oben) und falsch positive Detektionen (unten) in den Klassen "Liegend eingerollt" (links) und "Liegend seitlich" (rechts).

Der Kamerawinkel beeinflusst auch die Detektionen in der Klasse „Liegend seitlich“. Stehende Kühe werden, vermutlich aufgrund der guten Sichtbarkeit der Beine, als „Liegend seitlich“ klassifiziert (vgl. Abbildung 12, unten rechts), des Öfteren auch als „Liegend gerade“ wie in Abbildung 13 rechts. Viele falsch negative Detektionen in der

Klasse „Liegend seitlich“ sind auf die falsche Klassifizierung von Zwischenformen zurückzuführen, wie oben rechts in Abbildung 12 dargestellt. Die Sicherheitswahrscheinlichkeit von 56% spricht hier dafür, dass die Liegeposition immerhin nicht als klassisches „Liegend gerade“ interpretiert wird.

Des Weiteren erkennt das trainierte Modell Kühe weniger gut, die nah beieinander liegen, oder schlecht beleuchtet sind (vgl. Abbildung 13, links) und fasst sie zu einer Detektion zusammen. So entstehen viele falsch negative Detektionen. Auch randständige Kühe werden häufig nicht erkannt, obwohl solche Tiere im Trainingsdatensatz immer einbezogen und klassifiziert wurden.

Die Validierungsdaten deckten einen ganzen Tag in der Box ab, inklusive der Melkzeiten. Zu den Melkzeiten war keine Kuh auf dem Video, was vom trainierten Modell auch ausnahmslos richtig erkannt wurde, es gab keine einzige falsch positive Detektion.

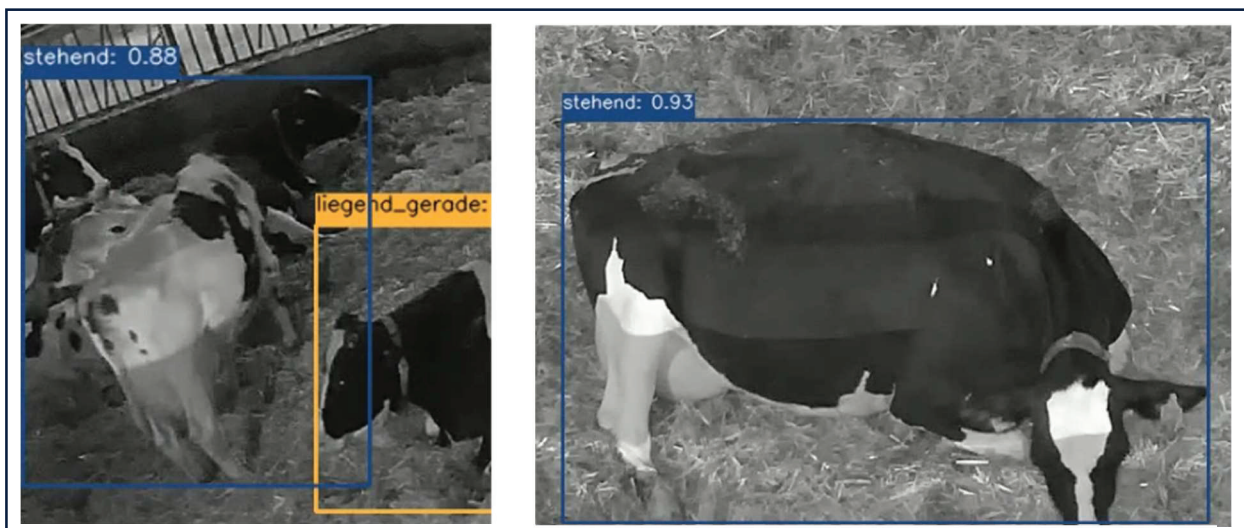


Abbildung 13: Fehlerbeispiel Validierung: eng stehende Kühe (links) und Kameraperspektive von oben (rechts).

Ansätze zur Erklärung des Leistungseinbruches und der Fehlertypen

Zur Erklärung des Leistungseinbruches in der Validierung und zur Erklärung der Fehlertypen müssen verschiedene Fehlerquellen betrachtet werden.

Als erster Ansatz zur Erklärung dieses Phänomens ist der Domain-Shift zu prüfen, also die Ähnlichkeit der Trainings- und Validierungsdaten. Das Modell ist darauf trainiert, die Körperhaltungen in einer bestimmten Umgebung zu detektieren. Eine stark abweichende Umgebung in den Validierungsdaten könnte zu schlechteren

Ergebnissen führen [7]. Das kann aber im vorliegenden Fall ausgeschlossen werden. Die Bilder, die in Training und Validierung verwendet wurden stammen aus demselben Versuchsaufbau. Die Detektionsumgebung ist also exakt dieselbe und verschiedene Aufnahmewinkel und Beleuchtungsverhältnisse finden sich im Trainingsdatensatz wieder.

Einen nächsten Einflussfaktor stellt die labelnde Person dar. Es ist möglich, dass der Trainings- und Validierungsdatensatz unterschiedlich streng bewertet wurden. Zur Validierung wurden einzelne Frames aus den Videos von Kamera 4 manuell detektiert. Durch die Bewegung ließen sich die Kühe und ihre Körperhaltungen besser erkennen, was dazu geführt haben könnte, dass einige Tiere und ihre Körperhaltung in den Validierungsdaten früher erkannt wurden als in den Trainingsdaten. Das betrifft vor allem Kühe am Rand des Bildes und Kühe in Zwischenformen von „Liegend gerade“ zu „Liegend seitlich“.

Einfluss des Trainingsdatensatzes

Wenn die Features der Klassen „Liegend eingerollt“ und „Liegend seitlich“ nicht zuverlässig erkannt werden, ist zu überprüfen, ob der Trainingsdatensatz zu klein war. In anderen Arbeiten zum Training eines neuronalen Netzwerkes wurden Datensätze in ähnlichen Größenordnungen verwendet. Jahn et. al verwendeten 2.224 Bilder für das Training von YOLOv8-pose [10]. Khin et. al trainierten ein YOLOv8m-cls-Modell mit 667 Bildern [20]. Bei Gong et. al umfasste der Trainingsdatensatz für YOLOv4 1800 Bilder [40]. Auf den Bildern in allen Datensätzen befand sich wenigstens eine Kuh, sodass die Anzahl der Labels wahrscheinlich höher ist als die Bilderzahl. Der Umfang des in dieser Arbeit genutzten Trainingsdatensatzes scheint im Vergleich dazu mit 1.350 Bildern und vielen Labels pro Bild akzeptabel. In zwei der oben erwähnten Arbeiten war die Verteilung der Klassen im Trainingsdatensatz deutlich unausgeglichen, trotzdem wurden gute Ergebnisse bei der Detektion erzielt, wie in dieser Arbeit auch. Allerdings erfolgte die Validierung nicht mit einem zusätzlich erstellten Validierungsdatensatz, sondern es wurden die Ergebnisse der internen Validierung während des Trainings verwendet [10, 20]. Ob die Modelle auf einen anderen Datensatz in derselben Umgebung dieselbe Detektionsleistung haben, kann nicht verglichen werden.

Aufgrund der guten Leistung des Modells in den Klassen „Stehend“ und „Liegend gerade“ kann angenommen werden, dass ein größerer Trainingsdatensatz speziell für die Klassen „Liegend eingerollt“ und „Liegend seitlich“ zu einer besseren Detektionsleistung führen würde. Die AP für „Liegend seitlich“ (91,9%) ist noch geringer als die für „Liegend eingerollt“ (93,1%), obwohl im Trainingsdatensatz mehr Labels für „Liegend seitlich“ vorhanden waren als für „Liegend eingerollt“. Das könnte auf die Zwischenformen zwischen „Liegend gerade“ und „Liegend seitlich“ zurückgeführt werden. Offenbar reichten die Beispiele im Trainingsdatensatz nicht aus, um die Features der beiden Klassen eindeutig voneinander abzugrenzen. Auch veränderte subjektive Bewertung in Training und Validierung könnte die Grenzen zwischen den Klassen aufgeweicht haben, was zu verrauschten Trainingsdaten geführt haben könnte [49]. Zur Verbesserung der Leistung des Modells in diesem Bereich könnte der Trainingsdatensatz vergrößert werden. Die Definition der Körperhaltung „Liegend seitlich“ könnte überarbeitet und präziser formuliert werden und es könnte eine „Eichung“ des Bewerter vor dem Labeln eingeführt werden, um den subjektiven Einfluss zu minimieren.

Nicht nur die Größe des Trainingsdatensatzes und der Label-Noise, sondern insbesondere seine Zusammensetzung beeinflusst die Leistungsfähigkeit des Modells nach dem Training. Mit steigendem Verhältnis von Mehrheits- zu Minderheitsklassen sinkt die Klassifikationsleistung in den Minderheitsklassen. Das Modell neigt dann dazu, einen Bias, also eine Tendenz zur Klassifikation der Mehrheitsklassen auszuprägen. Minderheitsklassen werden dann aus Prinzip seltener klassifiziert. Dieser Effekt ist stärker, je komplexer die Aufgabe ist [52]. Das könnte die Unterschiede in der Detektionsleistung zwischen den Klassen ebenfalls erklären. Um diesen Effekt auszugleichen wäre ein Oversampling der Trainingsdaten für die beiden Klassen möglich. Oversampling bedeutet, dass zufällige Trainings-Beispiele, hier für die Klassen „Liegend eingerollt“ und „Liegend seitlich“ kopiert werden, sodass der Anteil der Trainings-Beispiele dieser Klassen steigt. Es muss aber bedacht werden, dass Oversampling zu Overfitting führen kann [52].

Die Möglichkeit von Overfitting

Abgesehen davon muss auch Overfitting bei der Erklärung des Leistungsunterschiedes zwischen Training und externer Validierung bedacht werden. Die Daten für das Training und die interne Validierung stammen aus demselben Datensatz, der aufgeteilt wurde in 70% Trainings-, 20% interne Validierungs- und 10% Testdaten. Dementsprechend waren die Daten von Training und interner Validierung recht ähnlich. Kamera 4, deren Aufnahmen für die externe Validierung verwendet wurden ist nur zu 4% im Trainingsdatensatz vertreten (vgl. Tabelle 1). Es kann daher sein, dass die Perspektive und Beleuchtungsverhältnisse nicht exakt den Trainingsdaten entspricht. Wenn das Modell im Training die Hintergrund-Features der Labels zu genau gelernt hat, also Stallumgebung, Beleuchtung und Kameraposition mitgelernt hat, kann es die Features, die mit den Klassen verknüpft wurden, nicht auf den Validierungsdatensatz anwenden und die Detektionsleistung sinkt.

Für diese These spricht, dass sich ein typisches Symptom für Overfitting zeigt. Ein Modell erreicht im Training gute Ergebnisse, die Leistung ist bei der Detektion anderer Daten aber deutlich schlechter [43]. Der Recall bricht in der Validierung deutlicher ein als die Precision. Das heißt, dass das Modell nur noch eindeutige Fälle detektiert, Kühe in leicht abweichender Umgebung dagegen werden nicht erkannt. So entstehen viele falsch negative Detektionen. Dass die Klassen, für die es die wenigsten Trainingsbeispiele gab am schlechtesten generalisieren kann darauf hindeuten, dass das Modell in diesen Klassen am stärksten Hintergrund-Features einbezog.

Dagegen spricht, dass die Precision vergleichsweise höher bleibt als der Recall. Bei Overfitting ist es eher typisch, dass viele falsch Positive auftreten. Bei der Validierung traten dagegen viele falsch Negative auf. Das spricht eher für schlechte Trennbarkeit der Klassen als für Overfitting. Außerdem betrifft der Leistungseinbruch nur bestimmte Klassen. Der Hintergrund und die Beleuchtungsverhältnisse in beiden Datensätzen sollten sich auch nur wenig unterscheiden da der Versuchsaufbau derselbe ist.

6.4 Gesamtbewertung

Nach den Ergebnissen der Selbstüberprüfung scheint der Trainingsdatensatz wenig fehlerbehaftet. Die Trainingsergebnisse lassen darauf aufbauend ein Modell mit einer guten Detektionsleistung in allen Klassen und ohne offensichtliches Overfitting nach Auswertung der Loss-Kurven vermuten. Alle AP-Werte und die mAP liegt bei deutlich über 90%. In der Validierung ist die Leistung des trainierten Modells jedoch deutlich schlechter, vor allem in den Klassen „Liegend eingerollt“ und „Liegend seitlich“. Hauptgrund hierfür ist vermutlich ein unausgeglichener Trainingsdatensatz mit sehr wenig Trainingsbeispielen für zwei der vier Klassen. Diese Tatsache könnte für einen Bias zugunsten der Mehrheitsklassen gesorgt haben, zusätzlich könnte das Modell leicht overfittet sein, was an der Trainings-internen Validierung nicht abzulesen war.

In den Klassen „Stehend“ und „Liegend gerade“ werden die Kühe recht zuverlässig detektiert und klassifiziert. Fehler treten vor allem im Zusammenhang mit den Klassen „Liegend eingerollt“ und „Liegend seitlich“ sowie Zwischenformen, veränderten Winkeln und Perspektiven sowie bei randständigen Kühen auf (vgl. Abbildung 12, Abbildung 13). Vergleichsweise ist die Leistung des Modells im Training als hoch einzuschätzen, die in der Validierung als moderat [7, 10].

6.5 Limitationen der Arbeit

Die vorliegende Arbeit unterliegt inhaltlichen und methodischen Limitationen, die die Aussagekraft und Übertragbarkeit der Ergebnisse beschränken.

Die Auswirkungen des unausgewogenen Datensatzes wurden bereits diskutiert. Der ausgleichende Effekt des Nachlabelns war gering, könnte aber den Trainingsdatensatz verzerrt haben, weil viele schwarz-weiß-Aufnahmen enthalten waren. Um die Anzahl an Beispielen in den Klassen „Liegend eingerollt“ und „Liegend seitlich“ zu erhöhen, ist noch mehr Filterung des Bildmaterials nach diesen Körperhaltungen nötig. Des Weiteren waren die einzelnen Kameras und Aufnahmewinkel zu sehr unterschiedlichen Anteilen repräsentiert.

Das Vorlabeln durch eine untrainierte Version von YOLOv8 stellt eine Fehlerquelle dar, weil die Labels übersehen werden konnten und dann unbearbeitet in den Trainingsdatensatz einfließen. Die Genauigkeit der manuellen Labels war auch konzentrationsabhängig. Der subjektive Einfluss der annotierenden Person sowie der Unterschied beim Annotieren von Screenshots und Videos verringern die Vergleichbarkeit der beiden Datensätze. Auch die uneinheitliche Bewertung von Zwischenformen erhöht die Ungenauigkeit der Datensätze.

Die externe Validierung beschränkte sich auf die Gesamtzahl der Kühe in jeder Klasse und beinhaltete keine Einzeltierauswertung. Die Auswertung und tiefere Fehleranalysen sowie der Vergleich mit anderen Arbeiten sind dadurch limitiert. Die Verwendung einer festen IoU-Schwelle von 0,5 muss beim Vergleich dieser Arbeit mit anderen bedacht werden, weil die Ergebnisse bei dieser Bewertung optimistischer sind als bei anderen IoU-Schwellen.

Das trainierte Modell scheint Probleme zu haben, sich ähnelnde Klassen auseinanderzuhalten und scheint empfindlich gegenüber sich ändernden Winkeln und Perspektiven zu sein. Die Detektion wird stark von der Sichtbarkeit der Extremitäten beeinflusst. Bei stark verdeckten oder randständigen Objekten und durch die größere Anzahl von Kühen im Bild, die sich auch häufig verdecken, ist die Komplexität der Aufgabe erhöht. Durch den Leistungsabfall in der Validierung kann das trainierte Modell nur eingeschränkt auf andere Daten und Umgebungen angewendet werden.

6.6 Schlussfolgerung

Der Leistungseinbruch in der Validierung scheint multifaktoriell begründet. Als wahrscheinlichste Ursache erscheint eine Kombination aus unausgewogenem Trainingsdatensatz und Overfitting. Der Trainingsdatensatz könnte vergrößert werden. Es scheint notwendig, die Körperhaltungen, insbesondere „Liegend seitlich“ und „Liegend gerade“ klarer voneinander abzugrenzen und dadurch den Label-Noise zu verringern. Darüber hinaus sollten die Klassen ausgewogen im Trainingsdatensatz vertreten sein, um einem Bias des Modelles vorzubeugen. Zur Vermeidung von Overfitting könnten kleinere YOLOv8-Varianten mit dem verbesserten Datensatz trainiert werden oder das Training frühzeitig gestoppt werden, wenn eine ausreichende mAP erreicht ist. Weiterhin könnte das Training in mehrere Einheiten aufgeteilt werden. Zuerst könnte man gut sichtbare Kühe trainieren, dann Kühe in seltenen Körperhaltungen, verdeckte Kühe und randständige Kühe. So würde sich die Komplexität der Aufgabe nur langsam erhöhen und der Trainingsprozess liefere kontrollierter ab.

6.7 Ausblick

Die vorliegende Arbeit zeigt, dass das Computer-Vision- Modell YOLOv8 dazu in der Lage ist, Milchrinder in der Gruppe zu detektieren. Das Modell, dass in dieser Arbeit entstand, ist durch seine unausgeglichene Leistung in den verschiedenen Klassen nicht reif für eine Anwendung in der Praxis, zeigt aber gut, wo die Schwierigkeiten bei der Detektion in Gruppen liegen und worauf beim Training nachfolgender Modelle Wert gelegt werden sollte.

Modelle wie das aus der vorliegenden Arbeit können im Kontext des PLF genutzt werden, um eine dauerhafte Überwachung der Tiere zu realisieren und Trends bspw. in Milchviehherden sichtbar zu machen. Verschiedene Einsatzmöglichkeiten sind denkbar. Das Steh- und Liegeverhalten könnte ausgewertet werden, um Stress und Entspannung oder die Bequemlichkeit von Liegeflächen zu messen. Es könnte genutzt werden, um Kalbeverläufe zu überwachen. Wird die Fähigkeit der Neuronalen Netzwerke um Individuen-Erkennung ergänzt, könnten auch tierindividuelle Gesundheitsparameter oder Sozialverhalten überwacht werden. Technologien wie die

Halsband-Sensoren oder Pedometer könnten dadurch sogar überflüssig werden. Die Fähigkeit zur Detektion von Tieren in der Gruppe wäre dazu unerlässlich, da die meisten Rinder und andere landwirtschaftliche Nutztiere in Gruppen gehalten werden. Der Einsatz von Computer-Vision-Modellen auf landwirtschaftlichen Betrieben verspricht, die Tierüberwachung lückenlos, vom Personal unabhängig und objektiv zu gestalten.

Dadurch bietet der Einsatz von Computer-Vision-Modellen eine vielversprechende Möglichkeit, den heutigen Herausforderungen der Landwirtschaft zu begegnen. Die Tierüberwachung würde personalunabhängiger werden. Dadurch würde sich ein Personalmangel weniger auf die Tierüberwachung und das Tiergesundheitsmanagement auswirken. Die Arbeitszeit, die durch die Automatisierung des Prozesses gespart würde, könnte an anderer Stelle sinnvoll eingesetzt werden. Und letztendlich ist es auch vorstellbar, dass sich Computer-Vision-Programme mit herkömmlichen Herdenmanagement-Programmen verknüpfen lassen, sodass ein ganzheitliches System entsteht, das daten-gestützte Entscheidungen von der Kuh-individuellen Ebene bis hin zur Betriebswirtschaft ermöglicht.

7 Literaturverzeichnis

- [1] D. Berckmans, „General introduction to precision livestock farming,“ *Animal Frontiers*, Bd. 7, Nr. 1, 2017.
- [2] S. Neethirajan, „Artificial Intelligence and Sensor Innovations: Enhancing Livestock Welfare with a Human-Centric Approach,“ *Springer Nature*, Bd. 4, p. 77–92, 2023.
- [3] A. Stygar, Y. Gómez, G. Berteselli, E. Dalla Costa, E. Canali, J. Niemi, P. Llonch und M. Pastel, „A Systematic Review on Commercially Available and Validated Sensor Technologies for Welfare Assessment of Dairy Cattle,“ *Frontiers in Veterinary Science*, Bd. 8, Nr. 634338, 2021.
- [4] S. Ninomiya, Y. Goto, Huricha, H. Onishi, M. Kurachi und A. Ito, „Lying posture as a behavioural indicator of heat stress in dairy cows,“ *Applied Animal Behaviour Science*, Bd. 265, 2023.
- [5] C. Tucker, M. Bak Jensen, A. M. de Passillé, L. Hänninen und J. Rushen, „Invited review: Lying time and the welfare of dairy cows,“ *Journal of Dairy Science*, Bd. 104, Nr. 1, pp. 20-46, 2021.
- [6] K. Ito, M. A. G. von Keyserlingk, S. J. LeBlanc und D. M. Weary, „Lying behavior as an indicator of lameness in dairy cows,“ *Journal of Dairy Science*, Bd. 93, Nr. 8, pp. 3553-3560, 2010.
- [7] A. Bhujel, Y. Wang, Y. Lu, D. Morris und M. Dangol, „A systematic survey of public computer vision datasets for precision livestock farming,“ *Computers and Electronics in Agriculture*, Bd. 229, 2024.

- [8] M. Islam, G. Chen und S. Jin, „An Overview of Neural Network,“ *American Journal of Neural Networks and Applications*, Bd. 5, Nr. 1, 2019.
- [9] Ultralytics, „ultralytics.com,“ [Online]. Available: <https://docs.ultralytics.com/models/yolov8>. [Zugriff am 18 02 2026].
- [10] S. Jahn, S. Düpjan, V. Röttgen, J.-N. Jordt, S. Albrecht und T. Homeier-Bachmann, „Integrating YOLOv8-pose for localization and behavior tracking of calves in precision livestock farming,“ *Frontiers in Animal Science*, 16 01 2026.
- [11] G. Schmidt, „Tiergesundheitsüberwachung von Rindern via Videoüberwachung und Objektdetektion,“ Hochschule Neubrandenburg, 2023.
- [12] S. S. Nielsen, J. Alvarez und D. J. Bicout, „Welfare of dairy cows,“ *efsa Journal*, 2023.
- [13] C. Medrano-Galarza, J. Gibbons, S. Wagner, A. M. de Passilé und J. Rushen, „Behavioral changes in dairy cows with mastitis,“ *Journal of Dairy Science*, Bd. 95, Nr. 12, 2012.
- [14] V. Magierski, D. Frieten, J. Brinkmann, S. March, N. Heil und M. Koch, „Haltung Rind: Methodenhandbuch Nationales Tierwohl-Monitoring,“ Konsortium des Projektes Nationales Tierwohl-Monitoring (NaTiMon), 2024.
- [15] Nationales Tierwohlmonitoring, „nationales-tierwohl-monitoring.de,“ Johann-Heinrich-von-Thünen-Institut, 2026. [Online]. Available: <https://www.nationales-tierwohl-monitoring.de/projekt/tierwohl-definition?utm>. [Zugriff am 20 04 2026].
- [16] E. van Erp-van der Kooij, O. Almalik, D. Cavestany, J. Roelofs und F. van Eerdenburg, „Lying Postures of Dairy Cows in Cubicles and on Pasture,“ *animals*, Bd. 9, Nr. 4, 2019.

- [17] Niedersächsisches Ministerium für Ernährung, Landwirtschaft und Verbraucherschutz, „Tierschutzleitlinie für die Mastrinderhaltung,“ Hannover, 2018.
- [18] G. Charlton, D. B. Haley, J. Rushen und A. M. de Paissilé, *Journal of Dairy Science*, Bd. 97, Nr. 5, 2014.
- [19] A. D. Fisher, M. Stewart, G. A. Verkerk, C. J. Morrow und L. R. Matthews, „The effects of surface type on lying behaviour and stress responses of dairy cows during periodic weather-induced removal from pasture,“ *Applied Animal Behaviour Science*, Bd. 81, Nr. 1, pp. 1-11, 2003.
- [20] M. P. Khin, P. Tin, Y. Horii und T. T. Zin, „Predictive modeling of cattle calving time emphasizing abnormal and normal cases by using posture analysis,“ *Scientific Reports*, Bd. 14, Nr. 31871, 2024.
- [21] P. P. Nielsen, „Automatic registration of grazing behaviour in dairy cows using 3D activity loggers,“ *Applied Animal Behaviour Science*, Bd. 148, pp. 179-184, 2013.
- [22] I. Merkelyte, A. Siukscius und R. Nainiene, „The Role of Sensor Technologies in Estrus Detection in Beef Cattle: A Review of Current Applications,“ *Animals*, Bd. 15, Nr. 15, 2025.
- [23] L. N. Grinter, M. R. Campler und J. H. C. Costa, „Technical note: Validation of a behavior-monitoring collar's precision and accuracy to measure rumination, feeding, and resting time of lactating dairy cows,“ *Journal of Dairy Science*, 2019.
- [24] G. Vakulya, E. Hajnal, P. Udvardy und G. Simon , „In-Depth Development of a Versatile Rumen Bolus Sensor for Dairy Cattle,“ *Sensors*, Bd. 24, Nr. 21, 2024.

- [25] J. Pfrombeck, „Toward a systems-oriented evaluation of data-driven agriculture – the case of wearable sensors in dairy farming,“ Technische Universität München, 2023.
- [26] B. Pichlbauer, J. M. C. Gonzalez, M. Bobal, C. Guse, M. Iwersen und M. Drillich, „Evaluation of Different Sensor Systems for Classifying the Behavior of Dairy Cows on Pasture,“ *Sensors*, Bd. 24, Nr. 23, 2024.
- [27] S. Barney, S. Dlay, A. Crowe, I. Kyriazakis und M. Leach, „Deep learning pose estimation for multi-cattle lameness detection,“ *Scientific Reports*, Bd. 13, Nr. 4499, 2023.
- [28] C. Orhei, S. Vert, M. Mocofan und R. Vasiu , „End-To-End Computer Vision Framework: An Open-Source Platform for Research and Education,“ *Sensors (Basel)*, Bd. 21, Nr. 11, 2021.
- [29] Cattle Eye, „cattleeye.com,“ Cattle Eye Ltd., [Online]. Available: <https://cattleeye.com/>. [Zugriff am 05 03 2026].
- [30] Nedap Livestock Management, „nedap-livestockmanagement.com,“ [Online]. Available: <https://nedap-livestockmanagement.com/smartsight/>. [Zugriff am 05 03 2026].
- [31] Livestock 3D SA, „livestock3d.com,“ [Online]. Available: <https://livestock3d.com/>. [Zugriff am 05 03 2026].
- [32] A. Anagnostopoulos, B. E. Griffiths, N. Siachos und J. M. Neary, „Initial validation of an intelligent video surveillance system for automatic detection of dairy cattle lameness,“ *Frontiers in Animal Science*, 2023.
- [33] F. Altenberger und K. Lenz, „A Non-Technical Survey on Deep Convolutional Neural Network Architectures,“ arXiv.org, 2018.

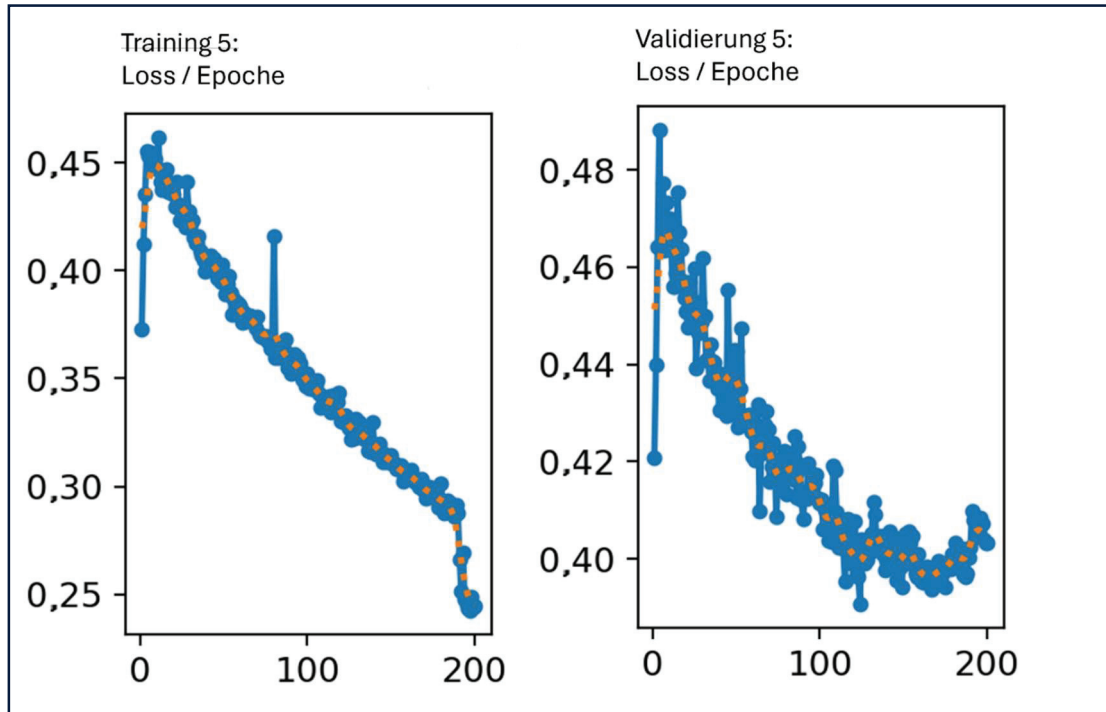
- [34] J. Schmidhuber, „Deep Learning in Neural Networks: An Overview,“ *Neural Networks*, Bd. 61, pp. 85-117, 2015.
- [35] J. Redmon, S. Divvala, R. Girshick und A. Farhadi, „You Only Look Once: Unified, Real-Time Object Detection,“ *arXiv:1506.02640*, 2016.
- [36] B. Ma, J. Zhao, S. Zhou, H. Wang, J. Xu, B. Liu, J. Hou und W. Liu, „Research on an Improved YOLOv8 Detection Method for Surface,“ *Micromachines*, Bd. 16, Nr. 12, 2025.
- [37] E. Malahina, M. Saitakela, S. Bulan, M. Lamabelawa und Y. Belutowe, „Teachable Machine: Optimization of Herbal Plant Image Classification Based on Epoch Value, Batch Size and Learning Rate,“ *Journal of Applied Data Sciences*, Bd. 5, Nr. 2, pp. 532-545, Mai 2024.
- [38] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid und S. Savarese, „Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression,“ 15 04 2019.
- [39] K. Oksuz, B. Can Cam, E. Akbas und S. Kalkan, „Localization Recall Precision (LRP): A New Performance Metric for Object Detection,“ 05 07 2018.
- [40] C. Gong, Y. Zhang, Y. Wei, X. Du, L. Su und Z. Weng, „Multicow pose estimation based on keypoint extraction,“ *plos one*, Bd. 17, Nr. 6, 03 Juni 2022.
- [41] A. M. Fahmy, „Confusion Matrix in Binary Classification Problems: A Step-by-Step Tutorial,“ *Journal of Engineering Research*, Bd. 6, Nr. 5, 2022.
- [42] Y. Jiang und R. Dale, „Mapping the learning curves of deep learning networks,“ *PLOS COMPUTATIONAL SCIENCE*, Bd. 21, Nr. 2, 2025.

- [43] C. Aliferis und G. Simon, „Overfitting, Underfitting and General Model Overconfidence and Under-Performance Pitfalls and Best Practices in Machine Learning and AI,“ in *Artificial Intelligence and Machine Learning in Health Care and Medical Sciences: Best Practices and Pitfalls*, Cham (CH), Springer, 2024.
- [44] Jupyter, „jupyter.org,“ [Online]. Available: <https://jupyter.org/about>. [Zugriff am 18 02 2026].
- [45] Ultralytics, „ultralytics.com,“ [Online]. Available: <https://docs.ultralytics.com/datasets/detect/coco/>. [Zugriff am 23 02 2026].
- [46] X. Fang, S. Hannuna und N. Campbell, „8-Calves Image dataset,“ *arxiv.org*, 2025.
- [47] I. A. Malek, B. Machani, A. M. Mevcha und N. H. Hyder, „Inter-observer reliability and intra-observer,“ *The Journal of Bone and Joint Surgery*, Bd. 88B, Nr. 9, pp. 1204 - 1206, 2006.
- [48] Z. Salekshahrezaee, J. L. Leevy und T. Khoshgoftaar, „A reconstruction error-based framework for label noise detection,“ *Journal of Big Data*, Bd. 8, Nr. 57, 2021.
- [49] Y.-W. Chen, Q. Song, X. Liu, P. S. Sastry und X. Hu, „On Robustness of Neural Architecture Search Under Label Noise,“ *Frontiers in Big Data*, Bd. 3, 2020.
- [50] L. B. Hunter, C. O'Connor, M. J. Haskell, F. M. Langford, J. R. Webster und K. J. Stafford, „Lying posture does not accurately indicate sleep stage in dairy cows,“ *Applied Animal Behaviour Science*, Bd. 242, 2021.
- [51] P. Song, „mljourney.com,“ [Online]. Available: <https://mljourney.com/diagnosing-model-overfitting-using-learning-curves/>. [Zugriff am 13 03 2026].

- [52] M. Buda, A. Maki und M. A. Mazurowski, „A systematic study of the class imbalance problem in convolutional neural networks,“ *Neural Networks*, Bd. 106, p. 249–259, 2018.
- [53] M. Alsaad, S. Dürr, D. Iten, W. Buescher und A. Steiner, „Locomotion behavior of dairy cows on traditional summer mountain farms in comparison with modern cubicle housing without access to pasture,“ *PLOS One*, 2022.

8 Anhang

Anhang 1: Loss-Kurve aus Training 5.



Verlauf der Loss-Kurve in Training 5, springender Kurvenverlauf.

Anhang 2: Ergebnisse der externen Validierung für jedes Video

00:00	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	53	7	2	0
False Negative	0	49	0	17
True Positive	605	1384	0	0
Precision	0,919453	0,994968	0	-
Recall	1	0,965806	-	0

00:20	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	15	106	0	1
False Negative	29	5	34	78
True Positive	1019	585	0	123
Precision	0,9855	0,8466	-	0,9919
Recall	0,9723	0,9915	0	0,6119

00:40	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	131	0	1	0
False Negative	0	142	0	119
True Positive	562	826	0	123
Precision	0,811	1	0	1
Recall	1	0,8533	-	0,5083

01:00	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	133	0	31	0
False Negative	0	191	0	143
True Positive	508	723	0	32
Precision	0,792512	1	0	1
Recall	1	0,791028	-	0,183

01:20	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	68	2	1	0
False Negative	5	261	21	125
True Positive	1082	536	3	1
Precision	0,9409	0,9963	0,75	1
Recall	0,9954	0,6725	0,125	0,0079

01:40	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	63	10	94	0
False Negative	10	128	0	118
True Positive	681	624	0	1
Precision	0,9153	0,9842	0	1
Recall	0,9855	0,8298	-	0,0084

02:00	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	75	0	0	2
False Negative	27	237	0	2
True Positive	1547	303	0	0
Precision	0,95	1,00	-	0
Recall	0,98	0,56	-	0

02:20	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	123	0	0	7
False Negative	0	140	5	45
True Positive	543	556	0	119
Precision	0,82	1	-	0,94
Recall	1	0,80	0	0,73

02:40	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	127	39	0	0
False Negative	39	33	132	125
True Positive	411	976	10	12
Precision	0,76	0,96	1	1
Recall	0,91	0,97	0,07	0,09

03:00	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	1	14	11	74
False Negative	265	56	16	0
True Positive	243	964	57	101
Precision	1,00	0,99	0,84	0,58
Recall	0,48	0,95	0,78	1

03:20	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	7	65	5	0
False Negative	144	3	0	55
True Positive	643	824	0	137
Precision	0,99	0,93	0	1
Recall	0,82	1,00	-	0,71

03:40	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	45	57	11	33
False Negative	14	45	105	3
True Positive	203	1437	85	108
Precision	0,82	0,96	0,89	0,77
Recall	0,94	0,97	0,45	0,97

04:00	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	0	0	10	1
False Negative	111	374	58	99
True Positive	732	1038	220	22
Precision	1	1	0,96	0,96
Recall	0,87	0,74	0,79	0,18

04:20	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	0	17	3	0
False Negative	104	32	19	229
True Positive	138	1652	121	246
Precision	1	0,99	0,98	1
Recall	0,57	0,98	0,86	0,52

04:40	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	8	275	0	3
False Negative	26	0	93	221
True Positive	325	1076	159	348
Precision	0,98	0,80	1	0,99
Recall	0,93	1	0,63	0,61

05:00	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	13	0	38	5
False Negative	8	365	11	5
True Positive	230	1687	79	148
Precision	0,95	1	0,68	0,97
Recall	0,97	0,82	0,88	0,97

05:20	Stehend	Lieg. gerade	Lieg. eingerollt	Lieg. seitlich
False Positive	3	6	34	10
False Negative	21	177	3	64
True Positive	375	1486	121	205
Precision	0,99	1,00	0,78	0,95
Recall	0,95	0,89	0,98	0,76

05:40	Stehend	Lieg. gerade	Lieg. eingerollt	Lieg. seitlich
False Positive	0	0	0	0
False Negative	0	0	0	0
True Positive	0	0	0	0
Precision	-	-	-	-
Recall	-	-	-	-

06:00	Stehend	Lieg. gerade	Lieg. eingerollt	Lieg. seitlich
False Positive	0	0	0	0
False Negative	0	0	0	0
True Positive	0	0	0	0
Precision	-	-	-	-
Recall	-	-	-	-

06:20	Stehend	Lieg. gerade	Lieg. eingerollt	Lieg. seitlich
False Positive	0	0	0	0
False Negative	0	0	0	0
True Positive	0	0	0	0
Precision	-	-	-	-
Recall	-	-	-	-

06:40	Stehend	Lieg. gerade	Lieg. eingerollt	Lieg. seitlich
False Positive	7	0	0	0
False Negative	167	0	0	0
True Positive	2011	0	0	0
Precision	1,00	-	-	-
Recall	0,92	-	-	-

07:00	Stehend	Lieg. gerade	Lieg. eingerollt	Lieg. seitlich
False Positive	0	0	0	0
False Negative	358	0	0	0
True Positive	3254	0	0	0
Precision	1	-	-	-
Recall	0,90	-	-	-

07:20	Stehend	Lieg. gerade	Lieg. eingerollt	Lieg. seitlich
False Positive	6	8	1	0
False Negative	204	7	0	0
True Positive	2387	190	0	0
Precision	1,00	0,96	0	-
Recall	0,92	0,96	-	-

07:40	Stehend	Lieg. gerade	Lieg. eingerollt	Lieg. seitlich
False Positive	29	0	9	1
False Negative	58	236	55	16
True Positive	1265	1023	1	0
Precision	0,98	1	0,1	0
Recall	0,96	0,81	0,02	0

08:00	Stehend	Lieg. gerade	Lieg. eingerollt	Lieg. seitlich
False Positive	1	7	8	0
False Negative	0	24	6	125
True Positive	363	1635	25	0
Precision	1,00	1,00	0,76	-
Recall	1	0,99	0,81	0

08:20	Stehend	Lieg. gerade	Lieg. eingerollt	Lieg. seitlich
False Positive	77	296	0	0
False Negative	4	3	208	149
True Positive	728	1415	35	5
Precision	0,90	0,83	1	1
Recall	0,99	1,00	0,14	0,03

08:40	Stehend	Lieg. gerade	Lieg. eingerollt	Lieg. seitlich
False Positive	104	74	20	0
False Negative	3	26	14	365
True Positive	1029	1356	27	37
Precision	0,91	0,95	0,57	1
Recall	1,00	0,98	0,66	0,09

09:00	Stehend	Lieg. gerade	Lieg. eingerollt	Lieg. seitlich
False Positive	9	33	13	5
False Negative	54	9	19	3
True Positive	1412	1231	12	118
Precision	0,99	0,97	0,48	0,96
Recall	0,96	0,99	0,39	0,98

09:20	Stehend	Lieg. gerade	Lieg. eingerollt	Lieg. seitlich
False Positive	1	0	0	1
False Negative	120	119	0	0
True Positive	1778	849	0	0
Precision	1,00	1	-	0
Recall	0,94	0,88	-	-

09:20	Stehend	Lieg. gerade	Lieg. eingerollt	Lieg. seitlich
False Positive	1	0	0	1
False Negative	120	119	0	0
True Positive	1778	849	0	0
Precision	1,00	1	-	0
Recall	0,94	0,88	-	-

09:40	Stehend	Lieg. gerade	Lieg. eingerollt	Lieg. seitlich
False Positive	0	1	3	1
False Negative	278	175	0	77
True Positive	1353	1150	0	44
Precision	1	1,00	0	0,98
Recall	0,83	0,87	-	0,36

10:00	Stehend	Lieg. gerade	Lieg. eingerollt	Lieg. seitlich
False Positive	25	45	0	0
False Negative	66	28	93	121
True Positive	973	1621	1	0
Precision	0,97	0,97	1	-
Recall	0,94	0,98	0,01	0

10:20	Stehend	Lieg. gerade	Lieg. eingerollt	Lieg. seitlich
False Positive	16	76	15	1
False Negative	5	25	3	130
True Positive	791	1399	6	15
Precision	0,98	0,95	0,29	0,94
Recall	0,99	0,98	0,67	0,10

10:40	Stehend	Lieg. gerade	Lieg. eingerollt	Lieg. seitlich
False Positive	51	4	0	1
False Negative	28	118	113	5
True Positive	1912	921	1	116
Precision	0,97	1,00	1	0,99
Recall	0,99	0,89	0,01	0,96

11:00	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	0	120	0	0
False Negative	390	16	73	121
True Positive	1240	800	0	0
Precision	1	0,87	-	-
Recall	0,76	0,98	0	0

11:20	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	67	17	0	12
False Negative	5	69	0	15
True Positive	1106	1521	0	0
Precision	0,94	0,99	-	0
Recall	1,00	0,96	-	0

11:40	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	94	1	14	1
False Negative	27	230	0	49
True Positive	1528	686	0	0
Precision	0,94	1,00	0	0
Recall	0,98	0,75	-	0

12:00	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	7	39	0	3
False Negative	11	51	24	94
True Positive	527	1005	97	38
Precision	0,99	0,96	1	0,93
Recall	0,98	0,95	0,80	0,29

12:20	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	12	0	18	2
False Negative	73	262	15	0
True Positive	1478	931	9	0
Precision	0,99	1	0,33	0
Recall	0,95	0,78	0,38	-

12:40	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	7	12	8	6
False Negative	30	41	53	54
True Positive	1007	1311	2	11
Precision	0,99	0,99	0,2	0,65
Recall	0,97	0,97	0,04	0,17

13:00	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	17	24	8	10
False Negative	69	34	0	19
True Positive	898	1045	0	66
Precision	0,98	0,98	0	0,87
Recall	0,93	0,97	-	0,78

13:20	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	12	38	0	15
False Negative	92	65	0	28
True Positive	1036	1345	121	79
Precision	0,99	0,97	1	0,84
Recall	0,92	0,95	1	0,74

13:40	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	8	13	11	2
False Negative	152	118	6	67
True Positive	1512	1015	2	31
Precision	0,99	0,99	0,15	0,94
Recall	0,91	0,90	0,25	0,32

14:00	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	5	38	4	8
False Negative	25	7	0	34
True Positive	1127	563	121	74
Precision	1,00	0,94	0,97	0,90
Recall	0,98	0,99	1	0,69

14:20	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	12	3	7	0
False Negative	162	72	6	0
True Positive	1881	438	10	0
Precision	0,99	0,99	0,59	-
Recall	0,92	0,86	0,63	-

14:40	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	0	46	28	0
False Negative	320	11	0	36
True Positive	1605	1033	0	0
Precision	1	0,96	0	-
Recall	0,83	0,99	-	0

15:00	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	5	16	2	3
False Negative	129	43	0	0
True Positive	1903	1103	0	0
Precision	1,00	0,99	0	0
Recall	0,94	0,96	-	-

15:20	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	61	4	2	10
False Negative	5	125	0	0
True Positive	414	1809	0	0
Precision	0,87	1,00	0	0
Recall	0,99	0,94	-	-

15:40	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	56	117	6	0
False Negative	0	6	0	216
True Positive	824	1210	7	13
Precision	0,94	0,91	0,54	1
Recall	1	1,00	1	0,06

16:00	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	0	0	0	0
False Negative	0	0	0	0
True Positive	0	0	0	0
Precision	-	-	-	-
Recall	-	-	-	-

16:20	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	0	1	0	0
False Negative	267	0	0	0
True Positive	1596	0	0	0
Precision	1	0	-	-
Recall	0,856683	-	-	-

16:40	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	0	5	0	0
False Negative	397	0	0	0
True Positive	1855	179	0	0
Precision	1	0,97	-	-
Recall	0,82	1	-	-

17:00	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	232	0	0	15
False Negative	0	501	98	6
True Positive	709	1575	9	32
Precision	0,75	1	1	0,68
Recall	1	0,76	0,08	0,84

17:00	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	232	0	0	15
False Negative	0	501	98	6
True Positive	709	1575	9	32
Precision	0,75	1	1	0,68
Recall	1	0,76	0,08	0,84

17:20	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	109	0	0	3
False Negative	0	304	127	79
True Positive	330	2098	5	122
Precision	0,75	1	1	0,98
Recall	1	0,87	0,04	0,61

17:40	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	109	0	0	3
False Negative	0	304	127	79
True Positive	330	2098	5	122
Precision	0,75	1	1	0,98
Recall	1	0,87	0,04	0,61

18:00	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	74	5	76	0
False Negative	4	186	10	266
True Positive	295	2067	55	83
Precision	0,80	1,00	0,42	1
Recall	0,99	0,92	0,85	0,24

18:20	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	35	0	0	14
False Negative	97	238	0	42
True Positive	1425	1053	0	82
Precision	0,98	1	-	0,85
Recall	0,94	0,82	-	0,66

18:40	Stehend	Lieg. gerade	Lieg. eingerollt	Lieg. seitlich
False Positive	44	1	0	10
False Negative	63	195	178	0
True Positive	941	1498	65	0
Precision	0,96	1,00	1	0
Recall	0,94	0,88	0,27	-

19:00	Stehend	Lieg. gerade	Lieg. eingerollt	Lieg. seitlich
False Positive	68	6	11	10
False Negative	23	75	5	6
True Positive	1072	1556	7	0
Precision	0,94	1,00	0,39	0
Recall	0,98	0,95	0,58	0

19:20	Stehend	Lieg. gerade	Lieg. eingerollt	Lieg. seitlich
False Positive	0	0	0	0
False Negative	221	345	0	0
True Positive	2228	901	0	0
Precision	1	1	-	-
Recall	0,91	0,72	-	-

19:40	Stehend	Lieg. gerade	Lieg. eingerollt	Lieg. seitlich
False Positive	9	0	32	1
False Negative	77	159	0	0
True Positive	2331	494	0	0
Precision	1,00	1	0	0
Recall	0,97	0,76	-	-

20:00	Stehend	Lieg. gerade	Lieg. eingerollt	Lieg. seitlich
False Positive	55	14	7	0
False Negative	39	200	6	83
True Positive	2550	560	36	2
Precision	0,98	0,98	0,84	1
Recall	0,98	0,74	0,86	0,02

20:20	Stehend	Lieg. gerade	Lieg. eingerollt	Lieg. seitlich
False Positive	62	0	0	29
False Negative	11	263	0	0
True Positive	1592	1017	0	0
Precision	0,96	1	-	0
Recall	0,99	0,79	-	-

20:40	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	203	0	2	4
False Negative	2	459	35	13
True Positive	1336	1639	139	0
Precision	0,87	1	0,99	0
Recall	1,00	0,78	0,80	0

21:00	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	89	0	7	4
False Negative	1	224	10	0
True Positive	1348	1575	118	0
Precision	0,94	1	0,94	0
Recall	1,00	0,88	0,92	-

21:20	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	32	5	1	57
False Negative	26	152	0	0
True Positive	1809	646	0	31
Precision	0,98	0,99	0	0,35
Recall	0,99	0,81	-	1

21:40	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	4	14	0	6
False Negative	154	54	15	0
True Positive	1242	770	106	0
Precision	1,00	0,98	1	0
Recall	0,89	0,93	0,88	-
22:00	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	4	0	26	25
False Negative	60	89	0	0
True Positive	799	1280	0	0
Precision	1,00	1	0	0
Recall	0,93	0,93	-	-

22:20	Stehend	Lieg. gerade	Lieg. engerollt	Lieg. seitlich
False Positive	0	0	0	4
False Negative	296	323	0	62
True Positive	850	1439	0	0
Precision	1	1	-	0
Recall	0,74	0,82	-	0

22:40	Stehend	Lieg. gerade	Lieg. eingerollt	Lieg. seitlich
False Positive	14	12	6	0
False Negative	144	113	0	25
True Positive	933	1216	0	14
Precision	0,99	0,99	0	1
Recall	0,87	0,91	-	0,36

23:00	Stehend	Lieg. gerade	Lieg. eingerollt	Lieg. seitlich
False Positive	0	129	3	15
False Negative	230	6	81	3
True Positive	823	1224	29	0
Precision	1,00	0,90	0,91	0
Recall	0,78	1,00	0,26	0,00

23:20	Stehend	Lieg. gerade	Lieg. eingerollt	Lieg. seitlich
False Positive	8	3	7	13
False Negative	13	60	1	0
True Positive	516	1523	53	0
Precision	0,98	1,00	0,88	0
Recall	0,98	0,96	0,98	-

23:40	Stehend	Lieg. gerade	Lieg. eingerollt	Lieg. seitlich
False Positive	0	49	0	0
False Negative	516	24	0	3
True Positive	2078	124	0	0
Precision	1,00	0,72	-	-
Recall	0,80	0,84	-	0,00

9 Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt.

Ort, Datum

Name