

Aus dem Department für Kleintiere und Pferde  
der Veterinärmedizinischen Universität Wien

Klinische Abteilung für Anästhesiologie und perioperative Intensivmedizin  
(Leiterin: Univ.-Prof. Dr.med.vet. Martina Mosing Dipl. ECVAA, MANZCVS)

**Analyse der Treffsicherheit von Annotierungen zwei verschiedener  
Video-tracking Programme vorbestimmter Körperteile bei  
hospitalisierten Pferden**

Diplomarbeit

Veterinärmedizinische Universität Wien

vorgelegt von  
Alessandra Woppel

Wien, im Oktober 2024

Betreuerin und 1. Begutachterin: Priv.-Doz. Dr.med.vet. Ulrike Auer

Klinische Abteilung für Anästhesiologie und  
perioperative Intensivmedizin

Department für Kleintiere und Pferde

Veterinärmedizinische Universität Wien

2. Begutachter:

Priv.-Doz. Dr. med. vet. Iwersen Michael

Klinische Abteilung für Bestandsbetreuung bei  
Wiederkäuern

Department für Nutztiere und öffentliches  
Gesundheitswesen in der Veterinärmedizin

Veterinärmedizinische Universität Wien

## INHALTSVERZEICHNIS

|        |                                                            |    |
|--------|------------------------------------------------------------|----|
| 1.     | EINLEITUNG .....                                           | 7  |
| 2.     | LITERATURÜBERSICHT .....                                   | 8  |
| 3.     | MATERIAL UND METHODIK .....                                | 12 |
| 3.1.   | MATERIAL .....                                             | 12 |
| 3.1.1. | Patienten.....                                             | 12 |
| 3.1.2. | Hardware .....                                             | 12 |
| 3.1.3. | Videoprogramm .....                                        | 12 |
| 3.2.   | METHODIK.....                                              | 13 |
| 3.2.1. | Bildmaterial .....                                         | 13 |
| 3.2.2. | Annotierung der Körperpunkte .....                         | 13 |
| 3.2.3. | "Key point prediction" der Bilder .....                    | 15 |
| 3.2.4. | Bestimmung der Genauigkeit der "key point" Erkennung ..... | 15 |
| 3.2.5. | Ergebniskategorien der ausgewerteten Körperpunkte .....    | 15 |
| 3.2.6. | Auswertung .....                                           | 16 |
| 3.2.7. | Berechnungen .....                                         | 17 |
| 3.2.8. | Statistik .....                                            | 17 |
| 4.     | ERGEBNISSE .....                                           | 18 |
| 4.1.   | AUSGEWERTETE BILDER .....                                  | 18 |
| 4.2.   | BERECHNUNGEN.....                                          | 20 |
| 4.3.   | BLAND ALTMAN .....                                         | 21 |
| 5.     | DISKUSSION .....                                           | 24 |
| 5.1.   | ANNOTATION MIT "COCO ANNOTATOR" .....                      | 24 |
| 5.2.   | VERGLEICH DER GENAUIGKEIT BEIDER PROGRAMME .....           | 25 |
| 6.     | ZUSAMMENFASSUNG .....                                      | 29 |

|     |                            |    |
|-----|----------------------------|----|
| 7.  | ABSTRACT .....             | 30 |
| 8.  | LITERATURVERZEICHNIS ..... | 31 |
| 9.  | ABBILDUNGSVERZEICHNIS..... | 34 |
| 10. | TABELLENVERZEICHNIS .....  | 35 |

## **Eigenständigkeitserklärung**

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig angefertigt, keine anderen als die angegebenen Hilfsmittel benutzt und alle aus ungedruckten Quellen, gedruckte Literatur oder aus dem Internet im Wortlaut oder im wesentlichen Inhalt übernommenen Formulierungen und Konzepte gemäß den Richtlinien wissenschaftlicher Arbeiten zitiert, durch Fußnoten gekennzeichnet bzw. mit genauer Quellenangabe kenntlich gemacht habe. In dieser Arbeit wurde ChatGPT von Open AI als Hilfe für die Übersetzung der Zusammenfassung von deutscher Sprache in die englische Sprache verwendet.

## **Abkürzungsverzeichnis**

|     |                              |
|-----|------------------------------|
| CNN | Convolutional neural network |
| DL  | Deep learning                |
| DNN | Deep neural network          |
| FN  | False negative               |
| FP  | False positive               |
| KI  | Künstliche Intelligenz       |
| TN  | True negative                |
| TP  | True positive                |
| VE  | Vorderextremität             |

## 1. Einleitung

Die Erkennung der Ausprägung von Schmerzen bei Pferden stellt seit vielen Jahren eine Herausforderung in der Pferdemedizin dar. Um eine objektivere Schmerzerkennung ermöglichen zu können, wurden Pain Scores erstellt, die Veränderungen der Gesichts- und Körperpositionen kategorisieren. Diese Art der Schmerzdeutung bietet mehrere Limitationen, eine davon ist die Anwesenheit der Kliniker, die Pferde Verhaltensänderungen verbergen lässt (1). Die eigene persönliche Einschätzung der Körperpositionen ist subjektiv und nicht einheitlich. Um diese Limitationen zu verbessern, werden neue Wege der Schmerzerkennung gesucht, vorherrschend sind dabei „image proccessing“ Programme, die künstliche Intelligenz verwenden. In den letzten Jahren wurden neue Techniken entwickelt, um Körperhaltungen von Menschen und Tieren schnell und präzise aufzeichnen zu können. Die schnelle Entwicklung von „Deep Learning“ Methoden führt zu einer rasanten Verbesserung in vielen Bereichen der Veterinärmedizin, Beispiele hierfür sind Forschungen in den Bereichen der Neurowissenschaften (2), des Stallmanagements (3) oder der Ganganalysen (4).

Die Vorteile einer automatisiert ausgewerteten Videoüberwachung sind dabei eine ungestörte Umgebung für das Tier, vereinfachte Sammlung des Datenmaterials und präzisere Ergebnisse. Eine kontinuierliche Überwachung mittels Videokamera nimmt dabei jegliche Körperpositionen und Verhaltensänderungen wahr und kann für die weitere Auswertung eine große Hilfe darstellen. Obwohl schon erste Ansätze vielversprechend sind, gibt es noch einige Herausforderungen bei dem Einsatz dieser Technologie.

Ziel dieser Diplomarbeit ist es, die zwei Videotracking Programme „Coco Annotator“ und „Loopy“ hinsichtlich ihrer Genauigkeit der Erkennung und Markierung vordefinierter Körperpunkte zu vergleichen. Als Hypothese wird angenommen, dass „Coco Annotator“ eine höhere Treffsicherheit bei der Erkennung von vordefinierten Körperpunkten als „Loopy“ aufweist.

## 2. Literaturübersicht

Der Einsatz von Videotracking Programmen findet immer häufigere Anwendung in der Forschung, vorrangig wird diese Technologie in der Verhaltensforschung (zum Beispiel Ethovision oder MiceProfiler (5)) und im Stallmanagement eingesetzt. Die ersten Videotracking Systeme sind in den frühen 1990er Jahren präsentiert worden, schnell haben sich dabei Vorteile, wie mehr Flexibilität und eine erhöhte Genauigkeit bei den Ergebnissen gezeigt. Während bei früheren Systemen noch mit dem Mauszeiger der Pfad des Tieres verfolgt werden musste, entwickelten sich schnell technisch ausgereifere Methoden (6). Beispiele für die Verwendung von Bildanalysen in Studien sind die Erfassung der Lokalisation und der Anzahl von Rindern auf Weidegrund (7) oder bei der Verhaltensforschung von sozialen und sexuellen Interaktionen von Mäusen (5).

Kontinuierliche Videoaufzeichnungen der Versuchsubstanzen ermöglichen es, auch nur schwach ausgeprägte Verhaltensänderungen zu erkennen. Der Einsatz von Videotracking Systemen trägt hier aktiv dazu bei, das Wohlbefinden der Tiere zu messen und Verbesserungsmaßnahmen durchführen zu können (8). Um beispielsweise aggressives Verhalten von Schweinen in Schweinebuchten schneller und genauer erkennen zu können, wurden Modelle erstellt, die aggressives von nicht aggressivem Verhalten mit hoher Genauigkeit in Videos erkennen können (9). Bei Pferden werden vermehrt Bewegungsanalysen und Verhaltensforschung mit Beschleunigungssensoren und dem Einsatz von „Convolutional Neural Networks“ (CNN) ausgewertet (10). Auch im Bereich der Landwirtschaft zeigen sich bedeutende Vorteile, Erkrankungen der Nutztiere werden schneller durch kontinuierliches Monitoring der Tiere erkannt und somit Produktionskosten gesenkt (11).

Die Dokumentation von Körperpositionen ist seit langer Zeit ein Interessensgebiet in der Veterinärmedizin (6). Mit der Digitalisierung war es möglich, Programme für die automatische Bildanalyse zu entwickeln.

Lange Zeit galt die Kombination von Videoaufnahmen mit gut erkennbaren und reflektierbaren Markern, die an verschiedenen Körperstellen der Testsubstanzen angebracht wurden, als der Goldstandard für die Messung von Bewegungen (12). Damit wurden erste Studien von Bewegungsabläufen durchgeführt und diese Technik zur Lahmheitsanalyse bei Hunden eingesetzt (13). Die Marker mussten vor den Aufnahmen platziert werden, die zu verfolgenden Bewegungen sind damit vor der Aufnahme bestimmt gewesen. Die Nachteile dieser Methode lag in ihrer Invasivität, Tiere neigen bei direkter Manipulation dazu, ihre gewohnten Abläufe

nicht einzuhalten. Die Messungen erfolgten meist in einem Laborbereich und der direkte menschliche Einfluss führt zu Verhaltensänderungen (12).

Unter markerloses Videotracking mithilfe einer Kamera versteht man die automatische Erkennung von Verhalten sowie Körperpositionen in Verbindung von Raum und Zeit, ohne dabei Personal einsetzen zu müssen. Tiere können in Echtzeit und über längere Zeitspannen beobachtet und gefilmt werden (15).

Künstliche Intelligenz ist die Fähigkeit von Computern menschliche Intelligenz nachzuahmen. Dieser Begriff wurde bereits 1956 von einer Gruppe von Informatikern erwähnt (14). Machine learning ist ein Teilbereich der künstlichen Intelligenz, „Deep learning“ ist wiederum ein Teilbereich des „Machine learnings“ (15). Vorbild für die Funktionsweise des „deep learnings“ ist das menschliche Gehirn, es bildet neuronale Netze. Convolutional Neural Networks ist eine Art von Deep learning Algorithmus, der vor allem für Bildanalysen von großen Datenmengen eingesetzt wird (16).

In den letzten zehn Jahren ist durch „deep learning“ (DL) ein großer Fortschritt gelungen. Im Jahr 2014 wurde das erste Paper publiziert, indem der Einsatz von DL bei 2D Bildern angewandt wurde (17). Von anfänglich 44% an korrekt gesetzten Körperpunkten, ist mittlerweile bei den meisten Videotracking Programmen eine Genauigkeit von zirka 94% möglich (2). Zahlreiche Videotracking Programme wie Ethovision, SocialScan oder VideoTrack verwenden 2D Videoaufnahmen, um Verhalten zu erforschen. Dazu werden hauptsächlich „Convolutional neural networks“ (CNNs) im Bereich der Analyse von Video- und Bilddateien zum Erkennen von visuellen Merkmalen eingesetzt. Ein neuronales Netzwerk nutzt dabei verschiedene Algorithmen (16). Der Einsatz von CNNs ist vielfältig, wie in Bereichen der Bildklassifikation, Objekterkennung oder auch das Erkennen von Körperpositionen (18). In einer Studie von Eerdeken et al. wurde zum Beispiel das Aktivitätenmuster mit Beschleunigungssensoren von Pferden getrackt und die Daten mithilfe von CNNs analysiert (10).

Für die automatisierte Verhaltensanalyse in Videos ist das Erkennen von Objekten, entweder in seiner Gesamtheit oder über markante Körperpunkte Voraussetzung (17). Diese automatisierte Analyse von Verhalten aus einem Video benötigt im Wesentlichen zwei Arbeitsschritte. Zuerst muss das Objekt im Video mittels z.B. CNN basierenden Erkennen von Körperteilen detektiert werden. Durch Bewegungsmuster eines Körperteils, beispielsweise des Kopfes, lässt sich mithilfe weiterer „Deep Learning“- Methoden auf das Verhalten des Tieres

schließen (19). So kann etwa das Verhalten "Fressen" anhand der Kopfbewegung automatisiert erkannt werden.

Für den ersten Arbeitsschritt, sollte die Wahl des Programms auf eine Software fallen, die mit gebräuchlichen Computern kompatibel ist (19). Je präziser ein Programm die Körperpunkte erkennen kann, desto besser kann in Folge das Verhalten und damit zum Beispiel die Schmerzerkennung in Videos klassifiziert werden (8). In dieser Arbeit sind die beiden Videotracking Programme „Loopy“ und „Coco Annotator“ verwendet worden. „Loopy“ ist ein webbasiertes Programm, mit dem man Videoanalysen durchführen kann. Beispiele hierfür sind die Aufzeichnung von Körperpositionen in Videos, Verhaltenscodierung oder die Setzung von Bildpunkten (20). „Coco Annotator“ ist ein Bildannotationstool, welches die Annotation von Bildern ermöglicht, die dann als Trainingsmaterial für DL Algorithmen verwendet werden (21).

Um in weiterer Folge die Tracking Modelle zur Verhaltenserkennung zu trainieren, sind genaue und präzise „ground truth“ Daten notwendig, deren Erstellung zeitintensiv ist und hoher Genauigkeit in der Erfassung bedarf (22). Dafür ist hochwertiges Ausgangsdatenmaterial notwendig (10). Die Bilder müssen eine hohe Varianz aufweisen. Je unterschiedlicher die Hintergründe, Lichtverhältnisse, die Fellfarbe von Tieren, die Kamerawinkel und andere Einflussfaktoren sind, desto weniger Datenmaterial und Zeitaufwand wird bei der manuellen Annotierung, sprich dem Setzen der Marker an vorbestimmten Körperstellen, benötigt (8).

In der Studie von Mathies et al. (2) im Jahr 2019 werden fünf Punkte für die Erstellung der Basisdaten beschrieben: 1.) Erstellung des Modells mit kleinen Datensätzen: es muss mit weniger als 20 000 Bildern erstellbar sein, 2.) die Präzision, die ein Programm erreicht, muss jenem Resultat entsprechen, welches Menschen durch Beschriftung erzielen können, 3.) das Programm muss robust gegenüber Änderungen des Versuchsaufbaus sein und auch bei einer Reanalyse zuverlässige Daten erbringen können, 4.) Tiere bewegen sich in 3D, es soll effiziente Lösungen in 3D geben, 5.) mehrere Individuen und Personen sollten erkannt werden können, um Versuche des sozialen Umfelds zu ermöglichen.

Für die Erstellung eines Modells zur Körperpunkterkennung wird das Datenset von Videobildern in ein Trainingsset mit meist zirka 25% aller Bilder aufgeteilt. Ein zweites Set dient der Überprüfung der Genauigkeit des Modells, wie gut beispielsweise Körperpunkte erkannt werden können. Ist die Performance des Modells nicht ausreichend, werden die Marker in diesen Bildern aus dem zweiten Set manuell an die richtige Position gesetzt. Die zusätzliche Information wird für ein neuerliches Training des Modells verwendet. Dieser Ablauf wird so

lange durchgespielt, bis eine ausreichend gute Präzision erreicht ist. Ein weiteres Set mit nicht für das Training verwendeten Bildern, dient anschließend für das Testen der finalen Performance des Modells (8).

### **3. Material und Methodik**

#### **3.1. Material**

##### **3.1.1. Patienten**

Für den Vergleich von zwei Videotracking Programmen sind Videos von sechs Pferden ausgewählt worden, diese sind in Boxen mit den Maßen von 3,5m x 3,5m untergebracht worden. Die Tiere haben ad libitum Wasser und 4x täglich Heu bekommen. Bei der Auswahl der Pferde wurde auf verschiedene Fellfarben, Abzeichen, Größe und Body Condition Score geachtet. Weiters wurden verschiedene Boxenhintergründe und unterschiedliche Lichtverhältnisse ausgewählt. Jeder Besitzer und jede Besitzerin haben eine Einverständniserklärung für die Videoüberwachung unterschrieben.

##### **3.1.2. Hardware**

Für die Aufnahme wurde die Kamera der Marke „GoPro (Inc, San Mateo, CA, USA)“ der Generation „Hero 4“ verwendet. Diese wurde auf einer Höhe von drei Metern an der Box angebracht und ist mit dem „Superwide“ Modus eingestellt worden. Gefilmt wurde jeweils über einen Zeitraum von drei Stunden. Um Datenmengen zu sparen, ist der „Timelapse“ Modus ausgewählt worden, dieser speichert zwei Bilder pro Sekunde ab. Die Auflösung ist dabei auf 2.7K eingestellt worden.

##### **3.1.3. Videoprogramm**

Für die Auswertung des Filmmaterials sind zwei Programme verglichen worden. Die Cloud basierte Software „Loopy“ (<http://loopb.io>, Loopbio Wien, Österreich) ist eine im Jahr 2015 von Max Hofbauer und John Stowers erstellte Software, die für Videoanalysen und Tierbeobachtungen eingesetzt werden kann (23).

Das Programm „Coco Annotator“ (Kurzfassung für “Common Objects in Context”, Justin Brooks) ist ein Videotracking Programm, welches auf Objekterkennung, Beschriftung von Bildern sowie Bewegungsanalysen spezialisiert ist (21).

### 3.2. Methodik

Für diese Diplomarbeit wurde ein Modell für die Erkennung von neun Körperpunkten bei Pferden, in Folge „key points“ genannt, mit dem Programm „Coco Annotator“ erstellt. Für das Programm „Loopy“ existierte bereits ein mit etwa 7200 Bildern trainiertes KI-Modell. Der Ablauf und die Ergebnisse sind von Kil et al. 2020 publiziert worden (1). Beide Modelle haben dieselben Videos analysiert und ausgewertet, danach wurde in dieser Diplomarbeit die Performance (Sensitivität, Genauigkeit und Fehlerrate) der beiden Programme miteinander verglichen.

In Abbildung 1 sind die Arbeitsschritte grafisch dargestellt, die für die Erstellung einer „key point prediction“ notwendig sind.

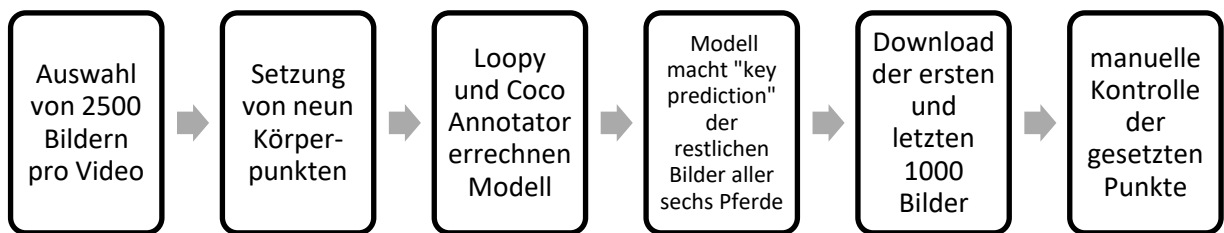


Abbildung 1: Arbeitsschritte für die Erstellung einer „key point“ prediction

#### 3.2.1. Bildmaterial

Für die Erstellung des Modells wurden Videos von sechs Pferden mit einer Länge von drei Stunden in das Programm „Coco Annotator“ hochgeladen. Das Programm wählte randomisiert 2500 Bilder aus zirka 120 000 Bildern aus. Bei Erstellung des Modells von Kil et al. in „Loopy“ wurden die zu annotierenden Bilder manuell ausgewählt (1).

#### 3.2.2. Annotierung der Körperpunkte

In jedem der 2500 Bilder wurden neun „key points“ manuell markiert, wobei jeder Punkt zur besseren Unterscheidung eine eigene Farbe zugewiesen bekommen hat. Als Körperpunkte wurden dabei die Nase, der Widerrist, der Schweif, das linke und rechte Ohr, die linke und rechte Vorderextremität sowie die linke und rechte Hinterextremität festgelegt. Die Definitionen für die richtige Lokalisation der einzelnen Körperpunkte, die bei „Loopy“ verwendet worden sind, wurden bei „Coco Annotator“ ebenfalls angewendet, um eine vergleichbare Annotierung zu gewährleisten.

Die Annotierung „Nase“ erfolgt am kaudalen Rand der Nüstern. Die „Ohren“ werden an den Ohrspitzen markiert. Der „Widerrist“, gebildet durch die Dornfortsätze der Brustwirbel, ist die höchste Stelle des Übergangs zwischen dem Hals und dem Pferderücken. Hier wird die höchste sichtbare Stelle annotiert. Die Markierung „Schweif“ ist am Schweifansatz zu annotieren. Ist dieser bedingt durch eine verdeckte Körperstellung nicht zu sehen, wird das Ende des Rückens annotiert. Die Vorder- und Hinterbeine werden am Saumbandansatz des Hufes annotiert (siehe Abbildung 2).

Nach fertiger Annotierung der 2500 Bilder wurde mit der Pixelinformation der neun „key points“, ein Modell zur automatischen Körperpunkt Erkennung in „Coco Annotator“ trainiert.



Abbildung 2: Annotierung mit dem Programm „Coco Annotator“ des Pferdes „Benebenuto“

### **3.2.3. "Key point prediction" der Bilder**

Für die Kontrolle, inwieweit die KI-Modelle die Körperpunkte korrekt erkennen, wurden die restlichen Bilder der sechs Videos einer „key point prediction“ mit dem Programm „Coco Annotator“ unterzogen. Dieselben Videos wurden in „Loopy“ mit dem bereits vorhandenen KI-Modell ebenfalls einer „key point prediction“ unterzogen.

### **3.2.4. Bestimmung der Genauigkeit der "key point" Erkennung**

Aus den sechs Videos, die der „key point prediction“ unterzogen worden sind, wurden jeweils die ersten und letzten 1000 Bilder ausgewählt. Insgesamt wurden in 12 000 Bildern jeweils fünf Körperpunkte ausgewertet, diese sind Nase, Widerrist, Schweif, linke Vorderextremität und rechte Vorderextremität. Nicht ausgewertet wurden Ohren und Hinterbeine, da die anderen Körperpunkte bereits ausreichend für die Analyse der korrekten Setzung der Körperpunkte im Vergleich beider Programme sind. Ohren sind primär für die Wahrnehmung der Umwelt und für die Gesichtsmimik verantwortlich, jedoch weniger für die Analyse des Bewegungsapparats. Die Hinterextremitäten sind aufgrund der häufig schlechten Sichtbarkeit ausgenommen worden.

Für die Zählung der gekennzeichneten „key points“ ist jedem Bild eine Zahl zwischen 0-2000 zugeordnet worden.

### **3.2.5. Ergebniskategorien der ausgewerteten Körperpunkte**

Für die Auswertung der Körperpunkte wurden verschiedene Ergebniskategorien definiert. Die zwei Hauptkategorien „Punkt richtig“ und „Punkt falsch“ sind jeweils in zwei Subkategorien unterteilt worden.

Die Setzung der Körperpunkte ist manuell durchgeführt worden, dabei wurde die Genauigkeit jedes Punktes visuell überprüft und die Entfernung zur anatomischen Zielstruktur beurteilt. Als korrekt gesetzt galt ein Punkt, wenn er den anatomischen Referenzbereich präzise traf, wie beispielsweise die Spitzen der Nasenlöcher oder markante Punkte wie zum Beispiel der Schweifansatz. Punkte wurden als fehlerhaft klassifiziert, wenn sie außerhalb des anatomischen Zielbereichs lagen, was auf eine falsche Erkennung oder Markierung hinwies.

Ein als richtig gewerteter Punkt bedeutet, dass das Modell die jeweiligen Körperpunkte an der richtigen Stelle erkannt hat, dies wird auch als „true positive“ bezeichnet. Als „true negative“ wird jener Zustand bezeichnet, bei dem das Programm korrekt den Körperpunkt nicht gekennzeichnet hat, da der Punkt nicht sichtbar gewesen ist.

Als „Punkt falsch“ wurde ein Ergebnis gewertet, wenn der Körperpunkt an einer falschen Stelle gekennzeichnet wurde („false positive“) oder keine Kennzeichnung erfolgte, obwohl der Punkt sichtbar gewesen wäre („false negative“).

Die nachstehende Grafik (siehe Abbildung 3) veranschaulicht die vorab geschilderten Kategorien.

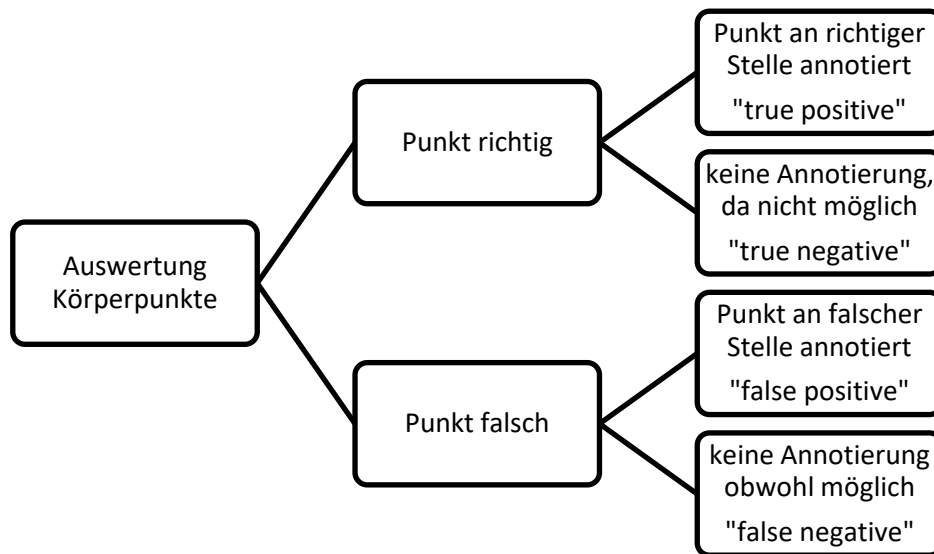


Abbildung 3: Auswertung Körperpunkte

### 3.2.6. Auswertung

Die nachstehende Tabelle 1 zeigt exemplarisch für das Pferd „Benebenuto“ die Ergebnisse, die mit dem Programm „Coco Annotator“ erreicht wurden. Die Anzahl der Bilder für „true positive“, „false positive“, „true negative“ und „false negative“ sind für jeden Körperpunkt angegeben.

Tabelle 1: Ergebnisse Benebenuto

|                  | Benebenuto    |                |               |                |
|------------------|---------------|----------------|---------------|----------------|
|                  | true positive | false positive | true negative | false negative |
| <b>Nase</b>      | 1900          | 16             | 38            | 35             |
| <b>Widerrist</b> | 1925          | 39             | 0             | 0              |
| <b>Schweif</b>   | 1982          | 18             | 0             | 0              |
| <b>Linke VE</b>  | 1946          | 49             | 4             | 1              |
| <b>Rechte VE</b> | 1936          | 57             | 2             | 0              |

### 3.2.7. Berechnungen

Um beide Programme vergleichen zu können, wurde jeweils die Sensitivität, die Fehlerrate („Error Rate“) und die Genauigkeit („Accuracy“) berechnet.

Die Sensitivität gibt Aufschluss darüber, wie groß der Anteil der als richtig positiv ausgewerteten Bilder in Bezug auf die Summe aller richtig und falsch positiv ausgewerteten Bilder ist. Je höher der Ergebniswert ist, desto mehr Bilder wurden vom Programm als richtig positiv erkannt.

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

Die „Error Rate“ zeigt das Verhältnis der als „false positive“ ausgewerteten Bilder an der Gesamtheit der als „true positive“ und „false positive“ Bilder an. Sie gibt an, wie hoch der Anteil der falsch positiv ausgewerteten Bilder an der Gesamtheit aller richtig und falsch ausgewerteten Bilder ist. Je höher der Ergebniswert ist, desto mehr Bilder wurden vom System als falsch positiv identifiziert.

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

Die „Accuracy“ zeigt das Verhältnis der Summe der als „true positive“ und „true negative“ bewerteten Bilder an der Gesamtheit aller Bilder. Je größer der Ergebniswert ist, desto genauer hat das Programm die Bilder korrekt ausgewertet.

$$\frac{TP + TN}{TP + FP + TN + FN}$$

### 3.2.8. Statistik

Die Ergebnisse sind als Summe pro „Key point“ angegeben. Um festzustellen, ob es einen signifikanten Unterschied zwischen den Ergebnissen beider Videoprogramme gibt, ist ein gepaarter T-Test pro Körperpunkt berechnet worden. Das Signifikanzniveau  $\alpha$  wurde mit 5% festgelegt.

Die Bland Altman Analyse wurde verwendet, um die Vergleichbarkeit beider Analyseverfahren von „Coco Annotator“ und „Loopy“ zu berechnen. Die Ergebnisse werden in grafischer Form als Punktdiagramm dargestellt.

## 4. Ergebnisse

### 4.1. Ausgewertete Bilder

Die nachstehenden Tabellen zeigen die Anzahl der Bilder pro Körperpunkt und Pferd, die vom jeweiligen KI-Programm als richtig bzw. falsch ausgewertet wurden. „Coco Annotator“ hat fast alle Körperpunkte als richtig erkannt und zeigt weniger falsch klassifizierte Bilder.

In Tabelle 2 wird die Anzahl der richtigen Bilder angezeigt, diese ergeben sich aus der Summe der „true positive“ und „true negative“ klassifizierten Bilder.

Tabelle 2: Ergebnisse richtige Bilder

| Anzahl der richtigen Bilder |            |      |           |         |          |           |
|-----------------------------|------------|------|-----------|---------|----------|-----------|
| Programm                    | Pferd      | Nase | Widerrist | Schweif | Linke VE | Rechte VE |
| <b>Coco Annotator</b>       | Benebenuto | 1949 | 1961      | 1982    | 1950     | 1943      |
|                             | Filine     | 1996 | 1738      | 1997    | 1376     | 1986      |
|                             | Lokko      | 1833 | 1836      | 1987    | 1993     | 1966      |
|                             | Lord       | 1973 | 1997      | 2000    | 1977     | 1979      |
|                             | Pascha     | 1672 | 1917      | 1981    | 1962     | 1786      |
|                             | Remy       | 1527 | 1973      | 1953    | 1876     | 1876      |
|                             |            |      |           |         |          |           |
| <b>Loopy</b>                | Benebenuto | 1933 | 1707      | 1910    | 798      | 1474      |
|                             | Filine     | 1606 | 197       | 1310    | 1155     | 476       |
|                             | Lokko      | 1583 | 1991      | 1983    | 1900     | 1896      |
|                             | Lord       | 1479 | 1888      | 1185    | 1163     | 1729      |
|                             | Pascha     | 1574 | 1380      | 511     | 1182     | 642       |
|                             | Remy       | 1151 | 1315      | 306     | 1288     | 1589      |

In der Tabelle 3 wird die Anzahl der falschen Bilder angezeigt, diese ergeben sich aus der Summe der „false positive“ und „false negative“ klassifizierten Bilder.

Tabelle 3: Ergebnisse falsche Bilder

|                       |            | Anzahl der falschen Bilder |           |         |          |           |
|-----------------------|------------|----------------------------|-----------|---------|----------|-----------|
| Programm              | Pferd      | Nase                       | Widerrist | Schweif | Linke VE | Rechte VE |
| <b>Coco Annotator</b> | Benebenuto | 51                         | 39        | 18      | 50       | 57        |
|                       | Filine     | 4                          | 262       | 3       | 624      | 14        |
|                       | Lokko      | 167                        | 164       | 13      | 7        | 34        |
|                       | Lord       | 27                         | 3         | 0       | 23       | 21        |
|                       | Pascha     | 328                        | 83        | 19      | 38       | 214       |
|                       | Remy       | 473                        | 27        | 47      | 124      | 124       |
| <b>Loopy</b>          | Benebenuto | 67                         | 293       | 90      | 1202     | 526       |
|                       | Filine     | 394                        | 1803      | 690     | 845      | 1524      |
|                       | Lokko      | 417                        | 9         | 17      | 100      | 104       |
|                       | Lord       | 521                        | 112       | 815     | 837      | 271       |
|                       | Pascha     | 426                        | 620       | 1489    | 818      | 1358      |
|                       | Remy       | 849                        | 685       | 1694    | 712      | 411       |

Das Programm „Coco Annotator“ hat signifikant ( $< 0,0001$ ) genauer die Punkte als „true positive“ und „false positive“ angezeigt im Vergleich zu „Loopy“.

In Tabelle 4 sind die Gesamtzahlen der ausgewerteten Bilder für jeden Körperpunkt ersichtlich. Die Anzahl der richtig ausgewerteten Bilder ist bei „Coco Annotator“ in jeder Kategorie höher als bei „Loopy“.

Tabelle 4: Gesamtanzahl der Bilder

|                       |          | Anzahl der Bilder |           |         |          |           |
|-----------------------|----------|-------------------|-----------|---------|----------|-----------|
| Programm              |          | Nase              | Widerrist | Schweif | Linke VE | Rechte VE |
| <b>Coco Annotator</b> | Richtig  | 10950             | 11422     | 11900   | 11134    | 11536     |
|                       | Falsch   | 1050              | 578       | 100     | 866      | 464       |
|                       | <b>Σ</b> | 12000             | 12000     | 12000   | 12000    | 12000     |
| <b>Loopy</b>          | Richtig  | 9326              | 8478      | 7205    | 7486     | 7806      |
|                       | Falsch   | 2674              | 3522      | 4795    | 4514     | 4194      |
|                       | <b>Σ</b> | 12000             | 12000     | 12000   | 12000    | 12000     |

## 4.2. Berechnungen

Die Sensitivität liegt bei „Coco Annotator“ bei allen Körperpunkten mit über 90% deutlich höher als bei „Loopy“. Die Fehlerrate ist bei „Loopy“ höher, lediglich der Körperpunkt „Schweif“ hat eine geringere „Error Rate“ als das Vergleichsprogramm. „Coco Annotator“ klassifiziert damit weniger Bilder falsch positiv als „Loopy“. Es zeigt sich eine deutlich bessere Genauigkeit („Accuracy“) bei „Coco Annotator“ als bei „Loopy“.

Tabelle 5: Berechnungen der Ergebnisse

| Programm              | Metriken    | %      |           |         |          |           |
|-----------------------|-------------|--------|-----------|---------|----------|-----------|
|                       |             | Nase   | Widerrist | Schweif | Linke VE | Rechte VE |
| <b>Coco Annotator</b> | Sensitivity | 92,54% | 97,42%    | 99,81%  | 98,55%   | 97,13%    |
|                       | Error Rate  | 1,99%  | 2,36%     | 0,64%   | 6,82%    | 1,90%     |
|                       | Accuracy    | 91,25% | 95,18%    | 99,17%  | 92,78%   | 96,13%    |
|                       |             |        |           |         |          |           |
| <b>Loopy</b>          | Sensitivity | 79,36% | 75,82%    | 60,11%  | 62,83%   | 59,95%    |
|                       | Error Rate  | 4,47%  | 8,96%     | 0,18%   | 17,27%   | 4,98%     |
|                       | Accuracy    | 77,72% | 70,65%    | 60,04%  | 62,38%   | 65,05%    |

### 4.3. Bland Altman

Die durchgeführte Bland Altman Analyse ergab einen signifikanten Unterschied zwischen den beiden Annotationsprogrammen.

Jedes Streudiagramm zeigt 30 Punkte, diese ergeben sich aus den fünf Körperpunkten der sechs Pferde. Jeder Punkt steht dabei für einen statistisch dargestellten Körperpunkt, der die Differenz gegen den Mittelwert abbildet. Die y-Achse stellt die Differenz der Ergebnisse dar, die x-Achse bildet den Mittelwert ab.

Der Mittelwert aller richtig annotierten Bilder beträgt bei „Coco Annotator“ 1738 Bilder, bei „Loopy“ sind es 1184 Bilder. Die Differenz der Mittelwerte beider Methoden sind 554 Bilder.

Der Korrelationskoeffizient für die richtig annotierten Bilder beträgt 0,39, dies deutet auf einen moderaten Zusammenhang der Messergebnisse hin.

Der Korrelationskoeffizient für die falsch annotierten Bilder liegt bei 0,18, dies spricht für einen geringen linearen Zusammenhang beider Werte.

#### Descriptive Statistics

| Variable   | Count | Mean     | Standard Deviation | 95.0% LCL of Mean | 95.0% UCL of Mean |
|------------|-------|----------|--------------------|-------------------|-------------------|
| C_FALSCH   | 30    | 101.9333 | 147.8681           | 46.71849          | 157.1482          |
| L_FALSCH   | 30    | 656.6334 | 517.0457           | 463.5653          | 849.7014          |
| Difference | 30    | -554.7   | 510.2517           | -745.2311         | -364.1689         |

#### Correlation

Correlation Coefficient = 0.188637

#### Abbildung 4: Ergebnisse der Bland Altman Analyse

Der Bland-Altman Plot zeigt die mittlere Differenz („Bias“) von 554 Bildern, dies zeigt einen großen Unterschied zwischen den Programmen.

Aus diesen Ergebnissen kann man schließen, dass die Methoden nicht miteinander vergleichbar sind.

### Bland-Altman Plot

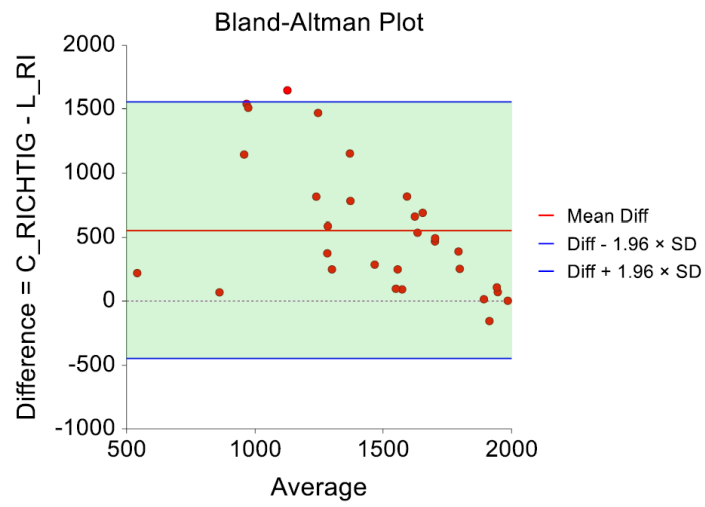


Abbildung 5: Bland-Altman Plot „richtig“

### Bland-Altman Plot

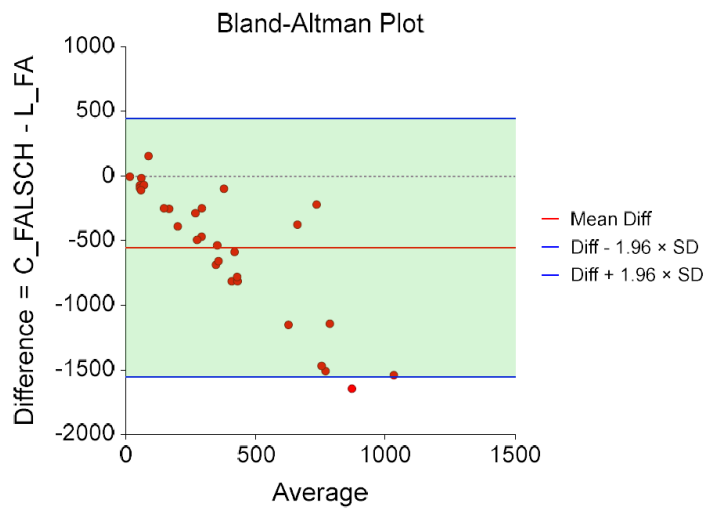


Abbildung 6: Bland-Altman Plot „falsch“

**Bland-Altman Analysis: Bias and Limits of Agreement for C\_RICHTIG and L\_RICHTIG**Limits of Agreement = Diff  $\pm$  1.96  $\times$  (Std Dev of Difference)

| Parameter                | Count | Value     | Standard Deviation | 95.0% LCL of Value | 95.0% UCL of Value |
|--------------------------|-------|-----------|--------------------|--------------------|--------------------|
| Bias (Difference)        | 30    | 554.7     | 510.2517           | 364.1689           | 745.2311           |
| Lower Limit of Agreement | 30    | -445.3934 | 161.0067           | -774.6891          | -116.0977          |
| Upper Limit of Agreement | 30    | 1554.793  | 161.0067           | 1225.498           | 1884.089           |

Abbildung 7: Berechnung der Bias und Limits of Agreement "richtig"

**Bland-Altman Analysis: Bias and Limits of Agreement for C\_FALSCH and L\_FALSCH**Limits of Agreement = Diff  $\pm$  1.96  $\times$  (Std Dev of Difference)

| Parameter                | Count | Value     | Standard Deviation | 95.0% LCL of Value | 95.0% UCL of Value |
|--------------------------|-------|-----------|--------------------|--------------------|--------------------|
| Bias (Difference)        | 30    | -554.7    | 510.2517           | -745.2311          | -364.1689          |
| Lower Limit of Agreement | 30    | -1554.793 | 161.0067           | -1884.089          | -1225.498          |
| Upper Limit of Agreement | 30    | 445.3934  | 161.0067           | 116.0977           | 774.6891           |

Abbildung 8: Berechnung der Bias und Limits of Agreement "falsch"

## 5. Diskussion

Das zentrale Ergebnis dieser Studie ist die Erkenntnis, dass „Coco Annotator“ die geeignetere Wahl darstellt, um präzise „Keypoints“ bzw. Markierungen an Pferden in Videos zu setzen.

### 5.1. Annotation mit "Coco Annotator"

Um eine exakte Annotation zu gewährleisten, sind Guidelines für die korrekte Setzung der jeweiligen Körperpunkte notwendig. Durch festgelegte Standards wird ein gleichbleibender Prozess bei der manuellen Setzung der Punkte eingehalten. Bei Beteiligung von mehreren Personen an einem Projekt, können neue Mitarbeiter schneller und effektiver eingearbeitet und Fehler vermieden werden. Ohne den Einsatz von solchen Guidelines, kann es zu großen Schwankungen der Setzung der Punkte kommen. Inkonsistente und abweichende Setzung der Körperpunkte führt zu einer geringeren Genauigkeit des Modells und falschen Ergebnissen. Abhängig von der Position der Pferde in der Box, konnten die einzelnen Körperpunkte besser oder schlechter gesetzt werden. Bei der manuellen Annotierung zeigte sich, dass bei seitlichen Aufnahmen alle Annotationspunkte gut gesetzt werden konnten.

Die Positionierung der Kamera über der Boxentür erzielte in den meisten Fällen gut auswertbare Bilder des gesamten Pferdekörpers bedingt durch die geringe Entfernung zu der Kamera. Dies zeigt, dass die Nähe der Kamera einen entscheidenden Einfluss auf die Qualität der aufgenommenen Bilder hat. Je weiter einzelne Tiere auf der Weide von der Kamera entfernt sind, desto weniger werden diese erkannt. Grund dafür ist die schlechter werdende Auflösung, dies ist abhängig von der Entfernung der aufgestellten Kamera. Bei der Studie von Benvenuti et al. konnte bei einer Distanz von 35 Metern eine ausreichend gute Auflösung festgestellt werden, bei 77 Metern erhöhte sich die Fehlerquote der Zählung der Tiere (24). Um das Verhalten des Pferdes in den Videos später besser beurteilen zu können, ist ein Winkel von 60° anstelle der Vogelperspektive gewählt worden. Dadurch konnten die meisten Körperpunkte gut gesehen werden. Der Schweif ist am öftesten verdeckt, dies ist auf die Position der Kamera auf der Boxentür zurückzuführen (1).

Die eingesetzte Kamera der Marke „GoPro“ zeigte mehrere Limitationen. Die zu dem Zeitpunkt verwendete Edition „Hero 4“ ist nicht die neueste Kamera gewesen, hier hätte eine aktuellere Version mit besseren technischen Daten verwendet werden können. Neuere Modelle haben eine höhere Auflösung und Lichtempfindlichkeit, welche qualitativ bessere Videos erstellen hätten können und damit das Setzen der Bildpunkte vereinfacht hätte.

Der Vorteil bei der Verwendung einer Actionkamera ist das geringe Gewicht und die rasche Entfernbarkeit aus der Box. Die Kamera ist rasch abgebaut und kann in einer beliebigen anderen Box schnell angebracht werden (8). Eine weitere Limitation stellen die Nachtstunden dar, in dieser Zeit ist mit der in der Studie gewählten Kamera keine Auswertung von den Videos möglich gewesen.

Die verwendete Software „Coco Annotator“ ist einfach in der Anwendung und sehr benutzerfreundlich. Die Bildpunkte können mit unterschiedlichen Farben markiert werden und sind deutlich zu erkennen. Erschwerend zeigte sich die korrekte Annotierung bei Bildern, bei denen Pferde von vorne aufgenommen wurden, liegende und in Bewegung befindliche Pferde. Auch Bilder mit teilweise verdeckten Körperteilen erwiesen sich als herausfordernd, da es oft schwer zu entscheiden war, ob das entsprechende Körperteil noch annotierbar ist oder nicht.

Bei Bildern im Tageslicht oder stark ausgeleuchteten Stalllicht war die manuelle Annotation leichter möglich als bei Bildern mit schlechten Lichtverhältnissen. In der Studie von Cirera et al. ist eine verringerte Anzahl an korrekt gesetzten Punkten bei zu starken Lichtverhältnissen beschrieben worden. Überbelichtete Bilder reduzieren ebenso wie unzureichend ausgeleuchtete Aufnahmen die Wahrscheinlichkeit einer präzisen Erkennung und Platzierung der „Key Body Points“ (8). Das Interieur und die Umgebung der Boxen trugen ebenfalls zu einer schwierigen Annotierung bei. Hohe Einstreu durch Stroh oder Pferdespäne machten die Annotierung der Pferdebeine äußerst schwierig, die Gitterstäbe der Boxen ermöglichen es den Pferden, die Köpfe durchzustrecken und erzeugten so nicht vollständig aufgenommene Pferdekörper. Ein weiterer Grund für nicht oder nicht vollständig sichtbare Körperteile war bei großen Heumengen zu Fütterungszeiten gegeben. In der Studie von Zhu et al. wurde ein Videotracking-Modell mit automatischer Erkennung von 17 Körperpunkten erstellt, wobei Hunde unterschiedlichster Rassen und Hintergründe als Grundlage dienten. Das HRNet-Modell erreichte dabei eine Genauigkeit von 80,9 % bei der Setzung der Körperpunkte. Die erhöhte Fehlerquote könnte auf die variierende Umgebung sowie die Vielfalt der Hunderassen zurückzuführen sein (25).

## **5.2. Vergleich der Genauigkeit beider Programme**

Die Genauigkeit der „key point“ Erkennung ist bei „Coco Annotator“ im Vergleich zu „Loopy“ schon nach dem ersten Training deutlich besser gewesen. Bei „Loopy“ wurde das Modell dreimal verbessert (Version 3) und erreichte eine „Accuracy“ mehr als 80% (1).

Ein unmittelbarer Vergleich der beiden Modelle in der Erkennung der „key points“ ist nur eingeschränkt möglich. Das Modell von „Loopy“ wurde nicht mit den gleichen Bildern erstellt wie das Modell von „Coco Annotator“. Für die Testung von „Coco Annotator“ sind dieselben Pferde bzw. Videos verwendet worden, die auch für das Training herangezogen wurden. Das bedeutet, die Bilder waren dem Trainingsmaterial ähnlich. Dies könnte eine mögliche Begründung für ein deutlich besseres Ergebnis von „Coco Annotator“ sein. Idealerweise hätte die Testung von „Coco Annotator“ mit neuen, unbekannten Pferdevideos stattfinden müssen. Es ist nicht bekannt welche KI-Technologie hinter den beiden Programmen steht und ob auf Basis dessen ein unmittelbarer Vergleich überhaupt reell ist.

Unabhängig davon gibt es einige Punkte, die zu Fehlern bei beiden Modellen führen können. Die unterschiedlichen Lichtverhältnisse erschwerten die Setzung der Körperpunkte bei beiden Videoprogrammen, insbesondere bei Dämmerungszuständen oder dunklen Boxen ohne Kunst- oder Tageslicht, ist die Bildqualität teilweise für eine korrekte Identifizierung der Körperpunkte und damit Platzierung der Marker nicht ausreichend gewesen. Bilder mit schlechter Bildqualität sind für die Erstellung eines Modells von hoher Bedeutung, da oftmals nicht optimale Lichtverhältnisse bei der Aufzeichnung gegeben sind. Ein robustes Tracking-System muss in der Lage sein, auch unter schwierigen Bedingungen wie niedriger Auflösung, Bewegungsunschärfe oder schlechten Lichtverhältnissen präzise Ergebnisse liefern zu können. Die Einbeziehung solcher Bilder in der Entwicklung und im Training des Modells erhöht dessen Genauigkeit und Anpassungsfähigkeit in realen Szenarien, in denen perfekte Bildqualität oft nicht gewährleistet ist.

Die unterschiedlichen Fellfarben und Abzeichen der gefilmten Pferde hatten je nach Tier mehr oder weniger Einfluss auf die erfolgreiche Setzung der Bildpunkte durch die Videoprogramme. Pferde mit vielen Abzeichen, insbesondere bei Nase und Widerrist, erschwerten für die Anwendungen die korrekte Setzung des Punktes, hier sind besonders hohe Fehlerquoten entstanden. Die Vorderbeine sind oft verwechselt worden, das linke Vorderbein wurde als das rechte gekennzeichnet und umgekehrt. Die Bewegung des Kopfes erschwerte ebenfalls eine richtige Körperpunktsetzung, in diesen Bildern lieferten die Programme oft keine Ergebnisse. Tendenziell ist die „key prediction“ bei Pferden in Boxen durch die eingeschränkte Bewegungsfreiheit etwas leichter möglich als bei Tieren in freier Wildbahn. Primaten zeigen bei uneingeschränkten Umgebungsbedingungen eine höhere Flexibilität der Gelenke und damit komplexere Gelenkpositionen, die wesentlich für die Verfolgung und Identifizierung der

Körperhaltung sind (25). Gut erprobt sind Videotracking Programme im Einsatzgebiet der Labortiermedizin bedingt durch gleichbleibende Umgebung.

Sehr dunkle Pferde wie Rappen oder Schwarzbraune haben einen geringen Kontrast bei sehr dunklen Bildern, hier sind die Grenzen der einzelnen Körperstellen sehr schwierig vorherzusagen. Die Ohrspitzen verschwimmen mit den Boxenwänden, der Widerrist ist nicht eindeutig abgrenzbar. Eine weitere Fehlerquelle stellte sich bei gefilmten Klinikpersonal dar, Personen, die in der Box aufgenommen wurden, erhöhten die Chance einer falschen Punktsetzung durch die Videoprogramme. Beispiele hierfür sind gelabelte Schuhe des Personals anstatt der Vorderbeine oder eine falsche Setzung des Punktes auf dem Futterkübel während der Futteraufnahme. Falsches Labelling stellt auch beim Videotracking mit mehreren Probanden eine potenzielle Fehlerquelle dar (2). Ein mögliches Problem kann die Unterscheidung des zu detektierbaren Tieres von Objekten wie Fahrräder, Zäune oder anderen Lebewesen sein. Ziegen haben sich schwer detektierbar vor Hecken gezeigt, Jungtiere sind teilweise unter Weidegras versteckt (3).

Bei den richtig ausgewerteten Bildern zeigte der Körperpunkt „Nase“ eine hohe Anzahl an richtigen Bildern, wobei „Coco Annotator“ bei allen Pferden höhere absolute Werte erzielte. Cirera et al. erwähnte ein schlechteres Ergebnis bei seinem erstellten Modell der Körperpunkte „Nasen“ und „Ohren“ verglichen mit den anderen Körperpunkten (8). Gründe für die unterschiedlichen Ergebnisse zwischen beiden Modellen könnten die unterschiedlichen Datensets sein. Ein Datenset könnte mehr Bilder mit verdeckten Nasen beinhaltet haben und so schlechtere Resultate im Training des Modells gezeigt haben. Bei dem Pferd „Filine“ mit dem Programm „Loopy“ gab es bei dem Körperpunkt „Widerrist“ eine stark verringerte Anzahl an richtigen Bildern. Ein Grund für das schlechte Ergebnis könnte die Liegeposition des Pferdes gewesen sein, „Loopy“ dürfte den Widerrist in dieser Position nicht erkannt haben. Der Körperpunkt „Schweif“ ist bei „Coco Annotator“ am besten erkannt worden, hier wurde der höchste Durchschnittswert erreicht. Bei „Lord“ wurden 100% der Bilder richtig gewertet, hier waren die Voraussetzungen der Bilder (gute Lichtverhältnisse, wenig Bewegung des Pferdes) am besten. Die „key prediction“ der „Vorderbeine“ hat bei beiden Programmen schlechtere Ergebnisse gebracht. Ein Grund dafür könnten die oft verdeckten Beine durch Heu und Einstreu gewesen sein. Auch war die Bildqualität der Vorderbeine“ etwas schlechter und damit schwieriger zu identifizieren.

Bei den falsch ausgewerteten Bildern gab es große Unterschiede zwischen den Videoprogrammen. Zusammenfassend ist zu sagen, dass „Loopy“ bei allen Körperpunkten

deutlich schlechter abgeschnitten hat als „Coco Annotator“, Ausreißer bei dem Programm „Loopy“ gab es bei den Punkten „Schweif“ und „rechte Vorderextremität“. Hier wurden besonders viele Punkte falsch gesetzt, mögliche Gründe dafür könnten sein, dass „Loopy“ bei schlechteren Lichtverhältnissen oder stärker bewegenden Pferdekörpern mehr Bilder als falsch erkennt. Ein anderer Grund könnte sein, dass zum richtigen Erkennen dieser Punkte ein weiteres Training mit noch mehr Bildern zu einer Verbesserung führen kann.

Die Grafiken der Bland-Altman Analyse zeigen die starken Differenzen der Mittelwerte der einzelnen Körperpunkte und unterstreichen die vorher diskutierten Ergebnisse.

## 6. Zusammenfassung

Ziel dieser Arbeit ist es, die zwei Videotracking Programme „Coco Annotator“ und „Loopy“ auf ihre Treffgenauigkeit der Setzung vordefinierter Körperpunkte zu vergleichen. Der Einsatz an künstlicher Intelligenz wird in vielen Teilen der Veterinärmedizin bereits regelmäßig angewendet, Beispiele dafür sind die automatisierte Erkennung von Körperpositionen oder im Bereich des „Pain Scorings“.

Für diese Diplomarbeit sind Videos von sechs hospitalisierten Pferden auf der Veterinärmedizinischen Universität Wien verwendet worden, die mithilfe eines Modells des jeweiligen Programms automatisch annotiert wurden. Nach Setzung der Körperpunkte durch KI-Modelle in den Videos, sind jeweils 2000 Bilder pro Pferd manuell überprüft worden und die Körperpunkte „Nase“, „Widerrist“, „Schweif“, „linke“ und „rechte Vorderextremität“ hinsichtlich mehrerer Kriterien überprüft worden. Die Punkte konnten dabei als „true positive“, „true negative“, „false positive“ und „false negative“ gelabelt werden. Die Auswertung der richtigen Bilder ergab eine höhere Trefferquote des Programms „Coco Annotator“ als bei „Loopy“, die Anzahl der falsch annotierten Bilder ist bei „Loopy“ höher gewesen. Die Bland-Altman Analyse ergab, dass es einen signifikanten Unterschied zwischen beiden Annotierungsprogrammen gibt.

Die Beurteilung des Datenmaterials zeigte mehrere Limitationen hinsichtlich der Bildqualität und der Umgebung, manche Pferde erschwerten eine korrekte Annotation durch ihre Fellfarbe oder Abzeichen. Körperpunkte wie die „Nase“ oder der „Schweif“ wurden bei beiden Programmen besser annotiert als der Widerrist. „Coco Annotator“ zeigte über alle Körperpunkte hinweg eine höhere Trefferquote an richtig annotierten Bildern als „Loopy“ und hat bei dieser Studie im Rahmen der Diplomarbeit besser abgeschnitten.

## **7. Abstract**

The aim of this thesis is to compare the accuracy of two video tracking programs, "Coco Annotator" and "Loopy," in placing predefined body points. Artificial intelligence is already regularly applied in various areas of veterinary medicine, such as automated recognition of body positions or pain scoring.

For this thesis, videos of six hospitalized horses at the University of Veterinary Medicine Vienna were used. These videos were automatically annotated using a model from each program. After the body points were set by the AI models in the videos, 2,000 images per horse were manually reviewed, and the body points nose, withers, tail, left and right forelimbs were evaluated based on several criteria. The points could be labelled as "true positive", "true negative", "false positive" and "false negative". The evaluation of the correct images showed a higher hit rate for the "Coco Annotator" program compared to "Loopy", with the number of incorrectly annotated images being higher in "Loopy". The Bland-Altman analysis revealed a significant difference between the two annotation programs.

The assessment of the data material revealed several limitations in terms of image quality and environment; some horses made accurate annotation more difficult due to their coat colour or markings. Body points such as the nose or tail were better annotated by both programs compared to the withers. Overall, "Coco Annotator" showed a higher hit rate for correctly annotated images across all body points than "Loopy" and performed better in this study within the framework of the thesis.

## 8. Literaturverzeichnis

1. Kil N, Ertelt K, Auer U. Development and Validation of an Automated Video Tracking Model for Stabled Horses. *Animals (Basel)* 2020; 10 (12). doi:10.3390/ani10122258
2. Mathis MW, Mathis A. Deep learning tools for the measurement of animal behavior in neuroscience. *Curr Opin Neurobiol* 2020; 60: 1–11. doi:10.1016/j.conb.2019.10.008
3. Bonneau M, Vayssade J-A, Troupe W, et al. Outdoor animal tracking combining neural network and time-lapse cameras. *Computers and Electronics in Agriculture* 2020; 168: 105150. doi:10.1016/j.compag.2019.105150
4. Eerdekens A, Deruyck M, Fontaine J, et al. Automatic equine activity detection by convolutional neural networks using accelerometer data. *Computers and Electronics in Agriculture* 2020; 168: 105139. doi:10.1016/j.compag.2019.105139
5. Matsumoto J, Urakawa S, Takamura Y, et al. A 3D-video-based computerized analysis of social and sexual interactions in rats. *PLoS One* 2013; 8 (10): e78460. doi:10.1371/journal.pone.0078460
6. Lucas P. J. J. Noldus, Andrew J. Spink, Ruud A. J. Tegelenbosch. *EthoVision: A versatile video tracking system for automation of behavioral experiments. Behavior Research Methods, Instruments, & Computers* 2001 (33): 398–414
7. Benvenuti MA, Coates TW, Imaz A, et al. The use of image analysis to determine the number and position of cattle at a water point. *Computers and Electronics in Agriculture* 2015; 118: 24–27. doi:10.1016/j.compag.2015.08.016
8. Martin-Cirera A, Nowak M, Norton T, et al. Comparison of Transformers with LSTM for classification of the behavioural time budget in horses based on video data. *Biosystems Engineering* 2024; 242: 154–168. doi:10.1016/j.biosystemseng.2024.04.014
9. Chen C, Zhu W, Steibel J, et al. Recognition of aggressive episodes of pigs based on convolutional neural network and long short-term memory. *Computers and Electronics in Agriculture* 2020; 169: 105166. doi:10.1016/j.compag.2019.105166
10. Eerdekens A, Deruyck M, Fontaine J, et al. Automatic equine activity detection by convolutional neural networks using accelerometer data. *Computers and Electronics in Agriculture* 2020; 168: 105139. doi:10.1016/j.compag.2019.105139

11. Neethirajan S. The role of sensors, big data and machine learning in modern animal farming. *Sensing and Bio-Sensing Research* 2020; 29: 100367.  
doi:10.1016/j.sbsr.2020.100367
12. Mathis A, Mamidanna P, Cury KM, et al. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat Neurosci* 2018; 21 (9): 1281–1289.  
doi:10.1038/s41593-018-0209-y
13. Yang-Chieh Fu MS, Bryan T. Torres, D. V.M., et al. Evaluation of a three-dimensional kinematic model for canine gait analysis. *American Journal of Veterinary Research* 2009 (Vol. 71, No 10): 1118–1122
14. Rene Y. Choi, Aaron S. Coyner, Jayashree Kalpathy-Cramer, Michael F. Chiang, J. Peter Campbell. Introduction to Machine Learning, Neural Networks, and Deep Learning (20.10.2024). Im Internet:  
<https://tvst.arvojournals.org/article.aspx?articleid=2762344>
15. Laurenz Wuttke. Deep Learning: Definition, Beispiele & Frameworks. Im Internet:  
<https://datasolut.com/was-ist-deep-learning/>; Stand: 12.12.2024
16. Intel Corporation. Convolutional Neural Networks (CNNs), Deep Learning und Computervision. Im Internet: <https://www.intel.de/content/www/de/de/internet-of-things/computer-vision/convolutional-neural-networks.html>; Stand: 26.08.2024
17. Alexander Toshev CS. DeepPose: Human Pose Estimation via Deep Neural Networks. *IEEE Conference on Computer Vision and Pattern Recognition* 2014: 1653–1660.  
doi:10.1109/CVPR.2014.214
18. J. Wang et al. Deep High-Resolution Representation Learning for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2020 (vol. 43, no. 10): pp. 3349-3364. doi:10.1109/TPAMI.2020.2983686
19. Bernardes RC, Lima M, Guedes R, et al. Ethoflow: Computer Vision and Artificial Intelligence-Based Software for Automatic Behavior Analysis. *Sensors* 2021 (21(9)).  
doi:10.3390/s21093237
20. Loopbio GmbH. Loopy. The complete solution for quantitative video analysis. Im Internet: <http://loopbio.com/loopy/>; Stand: 27.08.2024
21. COCO Annotator. Intro to Coco Annotation. Im Internet:  
<https://www.cocoannotator.com/#services>; Stand: 20.05.2024

22. Jacob M Graving, Daniel Chae, Hemal Naik, et al. DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. *eLife* 8:e47994 2020 Nov 30. doi:10.7554/eLife.47994
23. Max Hofbauer JS. Loopbio Products. Im Internet: <http://loopbio.com/index.html#products>; Stand: 20.05.2024
24. Benvenuti MA, Coates TW, Imaz A, et al. The use of image analysis to determine the number and position of cattle at a water point. *Computers and Electronics in Agriculture* 2015; 118: 24–27. doi:10.1016/j.compag.2015.08.016
25. Benjamin Y. Hayden, Hyun Soo Park, Jan Zimmermann. Automated Pose Estimation in Primates 2022. doi:10.1002/ajp.23348.

## 9. Abbildungsverzeichnis

|                                                                                           |    |
|-------------------------------------------------------------------------------------------|----|
| Abbildung 1: Arbeitsschritte für die Erstellung einer „key point“ prediction.....         | 13 |
| Abbildung 2: Annotierung mit dem Programm „Coco Annotator“ des Pferdes „Benebenuto“ ..... | 14 |
| Abbildung 3: Auswertung Körperpunkte.....                                                 | 16 |
| Abbildung 4: Ergebnisse der Bland Altman Analyse .....                                    | 21 |
| Abbildung 5: Bland-Altman Plot „richtig“ .....                                            | 22 |
| Abbildung 6: Bland-Altman Plot „falsch“ .....                                             | 22 |
| Abbildung 7: Berechnung der Bias und Limits of Agreement "richtig" .....                  | 23 |
| Abbildung 8: Berechnung der Bias und Limits of Agreement "falsch" .....                   | 23 |

**10. Tabellenverzeichnis**

|                                              |    |
|----------------------------------------------|----|
| Tabelle 1: Ergebnisse Benebenuto .....       | 16 |
| Tabelle 2: Ergebnisse richtige Bilder .....  | 18 |
| Tabelle 3: Ergebnisse falsche Bilder .....   | 19 |
| Tabelle 4: Gesamtanzahl der Bilder .....     | 19 |
| Tabelle 5: Berechnungen der Ergebnisse ..... | 20 |