TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

Diplomarbeit

Automatisierte Evaluierung der Ergonomie am Arbeitsplatz durch den Einsatz von Deep Learning

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines

Diplom-Ingenieurs

unter der Leitung von

Univ.-Prof. Dr.-Ing. Sebastian Schlund
(E330 Institute of Management Science,
Bereich: Human Centered Cyber Physical Production and Assembly Systems)

Dipl.-Ing. David Kostolani, BSc
(E330 Institute of Management Science,
Bereich: Human Centered Cyber Physical Production and Assembly Systems)

eingereicht an der TU Wien

Fakultät für Maschinenwesen und Betriebswissenschaften

von

Michael Wollendorfer

01526964 (066 482)

[redacted]

Pasching, im August 2021

Michael Wollendorfer

TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

Ich habe zur Kenntnis genommen, dass ich zur Drucklegung meiner Arbeit unter der Bezeichnung

Diplomarbeit

nur mit Bewilligung der Prüfungskommission berechtigt bin.

Ich erkläre hiermit Eides statt, dass ich meine Diplomarbeit nach den anerkannten Grundsätzen für wissenschaftliche Abhandlungen selbstständig ausgeführt habe und alle verwendeten Hilfsmittel, insbesondere die zugrunde gelegte Literatur, genannt habe.

Weiters erkläre ich, dass ich dieses Diplomarbeitsthema bisher weder im In- noch Ausland (einer Beurteilerin/einem Beurteiler zur Begutachtung) in irgendeiner Form als Prüfungsarbeit vorgelegt habe und dass diese Arbeit mit der von der begutachtenden Person beurteilten Arbeit übereinstimmt.

Pasching, im August 2021

Michael Wollendorfer

Danksagung

An dieser Stelle möchte ich mich bei all jenen bedanken, die mich bei der Erstellung dieser Diplomarbeit unterstützt und motiviert haben.

Zuerst gebührt mein Dank Herrn Univ.-Prof. Dr.-Ing. Sebastian Schlund und Herrn Dipl.-Ing. David Kostolani, BSc für die Betreuung und Zurverfügungstellung dieses spannenden Themas. Vor allem möchte ich Herrn Univ.-Prof. Dr.-Ing. Sebastian Schlund für die Kontaktherstellung mit der Wacker Neuson Linz GmbH danken. Des Weiteren gilt mein Dank Herrn Dipl.-Ing. David Kostolani, BSc für die gute Zusammenarbeit und die hilfreichen Anregungen bei der Erstellung dieser Arbeit.

Dank gebührt der Wacker Neuson Linz GmbH für die Unterstützung bei der Umsetzung der Diplomarbeit und die Zurverfügungstellung der notwendigen Ressourcen. Im Besonderen bedanke ich mich bei Herrn Dipl.-Ing. Martin Reitinger, der eine ausgezeichnete Zusammenarbeit mit dem Unternehmen ermöglichte und mir jederzeit mit Rat und Tat zur Seite stand.

Ein besonderer Dank gilt auch den beiden Produktionsmitarbeitern der Wacker Neuson Linz GmbH, die sich für den Feldversuch zur Verfügung stellten und sich während der Arbeit filmen ließen.

Abschließend möchte ich mich ganz herzlich bei meinen Eltern bedanken, die mich in all meinen Entscheidungen unterstützten und stets ein offenes Ohr für mich hatten. Insbesondere danke ich meinem Vater für die jahrelange finanzielle Unterstützung, welche mir dieses Studium ermöglichte und meiner Mutter für das Korrekturlesen meiner Diplomarbeit.

Ebenfalls möchte ich mich bei meinen Studienkollegen bedanken, die mir während des Studiums immer hilfsbereit zur Seite standen. Allen voran Herrn Markus Knöbl, mit dem ich nahezu jede Prüfung absolvierte.

Kurzfassung

ArbeitnehmerInnen in der Industrie sind oft schlechten Arbeitsbedingungen ausgesetzt. Dies führt zu einer erhöhten Rate an arbeitsbedingten Erkrankungen des Bewegungsapparates und beeinträchtigt dadurch die Produktivität [1]. Es existieren bereits einige Methoden für die Bewertung der Ergonomie am Arbeitsplatz. Jedoch stützen sich diese in der Praxis meist auf Beobachtungen durch Fachleute. Diese Analysen sind zeitaufwändig und erzielen subjektive Ergebnisse [2].

Daher wurde im Rahmen dieser Diplomarbeit eine Methode implementiert, welche die Ergonomie unter Verwendung von Computer-Vision-Algorithmen automatisch beurteilt. Für die Analyse ist lediglich die Aufzeichnung eines Videos des Arbeitsvorgangs mit einer handelsüblichen Kamera notwendig. Der Algorithmus verwendet ein künstliches neuronales Netzwerk zur Schätzung der menschlichen Pose. Dieses künstliche neuronale Netzwerk erkennt dabei die Personen in den einzelnen Bildern des Videos und gibt als Ergebnis die 2D- und 3D-Koordinaten der wichtigsten Körperpunkte aus. Basierend auf diesen Informationen wurden die Winkel zwischen einzelnen Körperteilen berechnet und daraus mithilfe der RULA-Methode ein Ergonomie-Score gebildet. Die Ergebnisse wurden in verschiedenen Heatmaps dargestellt, welche die Häufigkeit der Positionen einer Arbeitskraft, den Ergonomie-Score an den jeweiligen Positionen und die Häufigkeit der gefährlichen Körperhaltungen darstellen. Durch die Implementierung eines Tracking-Algorithmus ist es zusätzlich auch möglich, mehrere Arbeitskräfte zu analysieren oder unbeteiligte Personen zu ignorieren.

Der Algorithmus wurde im Rahmen eines Feldversuchs an zwei Arbeitsplätzen in der Produktion der Wacker Neuson Linz GmbH getestet und erzielte sehr gute Ergebnisse. Es konnten die Positionen aufgezeigt werden, an denen sich die Arbeitskraft am häufigsten aufhält. Des Weiteren ist in den Heatmaps sehr gut zu erkennen, an welcher Stelle eine schlechte Körperhaltung auftritt. Zusätzlich sind auch die Laufwege der Arbeitskraft ersichtlich.

Die automatische Analyse spart Kosten und Aufwand bei der Bewertung eines Arbeitsplatzes, ist objektiv und beeinträchtigt die Arbeitskraft nicht bei ihren Tätigkeiten. Die Darstellung der Ergebnisse in Form einer Heatmap fasst alle relevanten Informationen in einem Bild zusammen und sorgt für ein leichteres Verständnis.

Abstract

Industrial workers are often exposed to poor working conditions. This leads to an increased rate of work-related diseases of the musculoskeletal system and thereby it affects the productivity [1]. There are many methods for evaluating workplace ergonomics available, in practice, however, these methods are mostly based on observations by experts. Thus, ergonomic evaluation is time-consuming and can lead to subjective results [2].

So a method was implemented which automatically evaluates the ergonomics using computer vision algorithms. The sole requirement for the analysis is a video recording of the working process with a commercially available camera. The algorithm uses an artificial neural network to estimate the human pose. This artificial neural network recognizes the persons in each picture of the video and outputs as a result the 2D- and 3D-coordinates of the most important points of the body. Based on this information, the angles between the individual body parts were calculated and an ergonomics score was formed using the RULA method. The results were presented in various heatmaps, which show the frequency of the positions of the manual worker, the ergonomics score at the respective positions, and the frequency of dangerous postures. By implementing a tracking algorithm, it is also possible to analyze several workers or to ignore uninvolved persons.

The algorithm was tested as part of a field experiment at two workstations in the production of Wacker Neuson Linz GmbH and achieved very good results. The most frequent positions of the manual worker could be visualized. Furthermore, the heatmaps show very clearly where bad posture occurs. In addition, the routes taken by the worker are also apparent.

The automatic analysis saves costs and effort when evaluating a workplace, is objective, and does not interfere with the worker's activities. The presentation of the results in form of a heat map summarizes a great deal of information in one image and also ensures easier understanding.

Inhaltsverzeichnis

1 Einleitung

Dieses Kapitel gibt einen Überblick über das bearbeitete Themengebiet dieser Diplomarbeit und zeigt die aktuellen Probleme in dieser Hinsicht auf. Anschließend folgt eine kurze Beschreibung des Lösungswegs, welcher gemeinsam mit der Firma Wacker Neuson GmbH umgesetzt wurde.

1.1 Motivation

ArbeitnehmerInnen in der Industrie sind oft höheren körperlichen Anforderungen wie Überanstrengung, sich wiederholenden Bewegungen oder einer ungünstigen Körperhaltung ausgesetzt. Dies führt zu einer vergleichsweise hohen Rate arbeitsbedingter Erkrankungen des Bewegungsapparates und kann die Produktivität beeinträchtigen sowie die Produktionskosten erhöhen [1]. Eine ergonomische Analyse des Arbeitsplatzes verhindert in gewissem Maße arbeitsbedingte Erkrankungen des Bewegungsapparates und erhält bzw. erhöht die Produktivität, da sie ergonomische Risiken identifiziert und ineffiziente, unproduktive Bewegungen reduziert. Müdigkeit und mangelnde Motivation sind weitere ergonomische Faktoren, die zu einem Produktivitätsverlust führen können [2].

Eine quantitative Analyse der Ergonomie auf Montagearbeitsplätzen stützt sich derzeit oft nur auf Beobachtungen durch qualifizierte Fachleute. Obwohl eine direkte Beobachtung durch einen Menschen mit minimaler Störung der Arbeit und minimalem Geräteaufwand durchgeführt werden kann, ist die Analyse mit hohem Zeitaufwand verbunden und die Ergebnisse basieren auf einer subjektiven Bewertung. Somit besteht die Möglichkeit, dass verschiedene BeobachterInnen auch zu unterschiedlichen Ergebnissen kommen [2].

Zur Analyse der Ergonomie gibt es aber bereits unterschiedliche andere Methoden. Beispielsweise können Sensoren, die am menschlichen Körper angebracht sind, Beschleunigungen und Neigungen auf drei orthogonalen Achsen messen. Diese direkte Messung der Körperhaltung (z.B. unter Verwendung von Goniometern, Kraftsensoren, Beschleunigungsmessern und Elektromyographie) bietet zwar ein hohes Maß an Genauigkeit und liefert objektivere Informationen, kann jedoch durch experimentelle Kosten, Umgebung sowie technische und ethische Aspekte eingeschränkt sein [2] [3]. Eine andere Möglichkeit wäre, aktive oder passive Marker an bestimmten Teilen des menschlichen Körpers anzubringen. Eine oder mehrere Kameras erkennen dann die Position jedes Markers und somit die aktuelle Körperhaltung des Menschen [4] [5]. Beide Varianten haben aber den Nachteil, dass die überwachte Person einen unbequemen Anzug tragen muss, an dem die Sensoren und Marker montiert sind [2].

Es gibt auch bereits Methoden, für die keine besonderen Anzüge oder Sensoren am Körper notwendig sind. Sie verarbeiten aufgezeichnete Bilder mit Computer-Vision-Algorithmen, um die Bewegungen des menschlichen Körpers von der Hintergrundszene zu unterscheiden und seine Gelenke zu erkennen [6] [7]. Daraus lassen sich die Winkel zwischen den Gelenken ableiten und mit gängigen Analysemethoden wie z.B. dem Rapid Upper Limb Assessment (RULA) bewerten [8]. So wie die Beobachtung durch einen Menschen, ruft diese indirekte Messung mittels Kamera nur eine geringe Störung der Arbeit hervor. Im Gegensatz dazu liefert sie aber objektive Resultate und ist somit nicht mehr von der beobachtenden Person abhängig [2].

Heatmaps stellen in dieser Hinsicht eine Möglichkeit zur Visualisierung der menschlichen Bewegung dar. Sie werden bereits angewandt, um Konsumentenverhalten in Supermärkten [9] sowie an Webseiten [10] zu analysieren. Dementsprechend können sich Heatmaps auch zur Abbildung der Arbeitstätigkeiten eignen und stellen somit eine potenzielle Grundlage zur automatisierten qualitativen Analyse der Ergonomie am Arbeitsplatz dar. Beispielsweise beinhalten Heatmaps Informationen darüber, wo sich die Arbeitskraft bewegt und wonach sie am häufigsten greift. Eventuell zu lange Bewegungen können damit identifiziert und angepasst werden [11]. Des Weiteren ermöglicht diese Methode, nicht ergonomische Körperhaltungen an bestimmten Punkten darzustellen. Dadurch entsteht ein besseres Verständnis der Situation und es können gezieltere Maßnahmen entwickelt werden.

1.2 Problemstellung / Forschungsfragen

Eine ergonomische Verrichtung der Arbeitstätigkeiten wird heutzutage immer wichtiger. Durch den demografischen Wandel wird die arbeitende Bevölkerung immer älter und daher ist es wichtig, die Gesundheit der Menschen am Arbeitsplatz nicht zu schädigen. Eine schlechte Haltung bei der Durchführung von Tätigkeiten über einen längeren Zeitraum kann den Bewegungsapparat beeinträchtigen. Folgen daraus sind höhere Krankheitsausfälle und eine geringere Produktivität [1].

Es wurden bereits einige Methoden entwickelt, um die Ergonomie zu analysieren. Diese können in Selbstberichtsmethoden, Beobachtungsmethoden, direkte Messungen, indirekte Messungen und biomechanische Modelle eingeteilt werden [1]. Die Beurteilung der Ergonomie erfolgte in der Industrie bis jetzt meistens von beobachtenden Personen und führte dadurch zu subjektiven Ergebnissen bei relativ hohem Zeitaufwand. Eine kamerabasierte Analyse, welche zu den indirekten Methoden zählt, liefert hingegen objektive Ergebnisse und ist somit nicht mehr von einem Menschen abhängig. Zusätzlich kann sie automatisiert ausgeführt werden [2]. Diese Methoden sind allerdings zurzeit nur in der Forschung relevant. Daher sollte

diese Diplomarbeit den aktuellen Stand der Technik in diesem Themenfeld aufzeigen und einen Beitrag zu den indirekten Messungen leisten.

Daraus wurden folgende Forschungsfragen definiert:

- *Welche Methoden zur Analyse der Ergonomie mit Kameras gibt es bereits?*
- *Wie kann der Mensch in einem zweidimensionalen Video erkannt werden?*
- *Wie können die Bewegung einzelner Körperteile und die Ergonomie in Heatmaps dargestellt und interpretiert werden?*

1.3 Lösungsansatz / Arbeitspakete

Um einen Überblick über aktuelle kamerabasierte Methoden zur Analyse der Ergonomie zu geben, wurden aktuelle Entwicklungen im Rahmen einer Literaturrecherche aufgezeigt und gegenübergestellt.

Im nächsten Schritt wurde ein Programmieralgorithmus implementiert, der eine objektive Bewertung der Ergonomie eines Montagearbeitsplatzes unterstützen soll. Der Algorithmus erkennt den Mensch mithilfe eines Posenschätzers in einem Video des Montagevorganges und analysiert seine Körperhaltung. Der Posenschätzer liefert als Ausgabe die 2D- und 3D-Koordinaten der sogenannten Body-Keypoints. Zu den Body-Keypoints zählen zum Beispiel Schultern, Hände und Knie [12]. Aus diesen lassen sich die einzelnen Gelenkwinkel berechnen. Auf Basis der Methode zur Ergonomieanalyse RULA können die Gelenkwinkel dann mit einem Score bewertet werden.

Zur Darstellung der Ergebnisse dienen Heatmaps, in denen die Ergonomie-Scores bei der jeweiligen Position der Arbeitskraft aggregiert werden. Daraus können Verbesserungsmöglichkeiten hinsichtlich der Ergonomie am Arbeitsplatz abgeleitet werden. Beispielsweise ist es dadurch möglich, häufige unnötige oder belastende Bewegungen zu erkennen und daraus Maßnahmen zur Reduktion dieser Bewegungen zu entwickeln.

Für erste Tests während des Programmiervorgangs wurden frei zugängliche Bilder oder Videos aus dem Internet genutzt. Nach der Fertigstellung erfolgte ein Feldversuch an zwei realen Montage-Arbeitsplätzen der Firma Wacker Neuson Linz GmbH.

1.4 Aufbau und Struktur der Arbeit

Am Beginn der Arbeit werden die theoretischen Grundlagen der Haupt-Themengebiete beschrieben, welche zum besseren Verständnis der Diplomarbeit beitragen sollen. Ein wichtiges Themenfeld dieser Diplomarbeit ist die Ergonomie. Neben der Definition und der aktuellen Problematik der Ergonomie am Arbeitsplatz werden bereits existierende Analysemethoden vorgestellt. Das zweite große Themengebiet stellt die künstlichen

neuronalen Netzwerke dar, auf welchen die Schätzung der menschlichen Pose basiert. Dabei wird die grundsätzliche Funktionsweise beschrieben und insbesondere auf die faltenden neuronalen Netzwerke eingegangen, die eine effektive Bilderkennung ermöglichen. Abschließend werden im Grundlagen-Kapitel verschiedene Methoden der 2D- und 3D- sowie der Einzelpersonen- und Mehrpersonen-Posenschätzung erläutert.

Im Kapitel State-of-the-Art wird ein Überblick über aktuelle Entwicklungen der menschlichen Posenschätzung und der damit verbundenen Ergonomiebewertung gegeben. Es werden Methoden sowohl zur 2D- als auch zur 3D-Analyse vorgestellt. Anschließend folgen aktuelle Anwendungen der Posenschätzung für die Ergonomiebewertung. Zum Vergleich werden auch noch andere Methoden vorgestellt, welche die Ergonomie ohne Posenschätzung beurteilen.

Kapitel 4 beschreibt den erstellten Programmiercode. Dabei wird auf die Anforderungen der Programmierumgebung eingegangen und der verwendete Posenschätzer vorgestellt. Danach wird aufgezeigt, wie die Berechnung des Ergonomie-Scores aus den erhaltenen Keypoint-Koordinaten abläuft und wie die Darstellung in Heatmaps erfolgt. Zusätzlich wurde noch ein Tracking-Algorithmus erstellt, welcher ermöglicht, unbeteiligte Personen aus der Analyse herauszufiltern. Die Erklärung der Funktionsweise des Trackings schließt das Kapitel ab.

Im Anschluss wurde der erstellte Algorithmus in einem Feldversuch getestet. Dabei wurden zwei Arbeitsplätze in der Produktion der Wacker Neuson Linz GmbH analysiert. Das Kapitel 5 beschreibt den Ablauf sowie die verwendete Hardware und stellt die beiden Arbeitsplätze vor.

Kapitel 6 erläutert die Ergebnisse des Feldversuchs. Dabei werden die erzeugten Heatmaps und die daraus resultierenden Erkenntnisse beschrieben. Abschließend werden die Forschungsfragen aus Kapitel 1.2 beantwortet.

Das letzte Kapitel erläutert den gewonnenen Nutzen der Methode und zeigt die Einschränkungen auf. Zusätzlich werden die nächsten möglichen Schritte zur Weiterentwicklung vorgestellt, um den erstellten Algorithmus noch weiter zu verbessern und einen professionellen Einsatz zu ermöglichen.

2 Theoretische Grundlagen

Dieses Kapitel beschreibt die theoretischen Grundlagen dieser Diplomarbeit und soll so zum besseren Verständnis beitragen. Dabei werden die Themen „Ergonomie“, „künstliche neuronale Netzwerke“ und „Schätzung der menschlichen Pose“ näher vorgestellt.

2.1 Ergonomie

Dieser Abschnitt geht auf die Grundlagen und Ziele der Ergonomie ein und beschreibt die aktuellen Probleme in der Industrie. Anschließend werden die verschiedenen Methoden zur Erfassung der Ergonomie eines Produktionsarbeitsplatzes inklusive Beispielen vorgestellt.

2.1.1 Grundlagen

Der Begriff Ergonomie wurde in der Literatur erstmals im Jahr 1857 von Jastrzebowski definiert. Er kommt aus dem Altgriechischen und setzt sich aus den beiden Wörtern „ergon“ (deutsch: Arbeit, Werk) und „nomos“ (deutsch: Regel, Gesetz) zusammen [13].

Die internationale Fachgesellschaft für Ergonomie und Arbeitswissenschaft (IEA) definiert Ergonomie als „wissenschaftliche Disziplin, die sich mit dem Verständnis der Wechselwirkungen zwischen Menschen und anderen Elementen eines Systems befasst bzw. als der Berufsstand, der Theorie, Grundsätze, Daten und Verfahren auf die Gestaltung von Systemen anwendet, mit dem Ziel, das Wohlbefinden des Menschen und die Leistung des Gesamtsystems zu optimieren“ [14, p. 5].

Die Ziele der Ergonomie sind die Erleichterung der Ausführung einer Aufgabe sowie der Schutz und die Förderung der Sicherheit, Gesundheit und des Wohlbefindens einer Arbeitskraft. Dies kann durch die Optimierung von Aufgaben, Arbeitsmittel, Dienstleistungen und Umgebungen erreicht werden. Allgemein betrachtet, sollen sämtliche Elemente eines Systems an den Menschen angepasst werden, um effizientes und fehlerfreies Arbeiten sicherzustellen und die Menschen vor Gesundheitsschäden auch bei langfristiger Ausübung einer Tätigkeit zu schützen. Die Ergonomie hat sowohl wirtschaftliche und gesellschaftliche als auch umweltbezogene Auswirkungen. Die Anpassung eines Gestaltungsaspektes an die Bedürfnisse und Fähigkeiten des Menschen verringert beispielsweise die Fehleranfälligkeit, die Ablehnung des Systems und die Anzahl an Krankheitsausfällen [14].

Vor allem arbeitsbedingte Erkrankungen des Bewegungsapparates sind ein großes Problem in der Industrie. Zu diesen gehören Beeinträchtigungen der Körperstrukturen wie Muskeln, Gelenke, Sehnen, Bänder, Nerven, Knorpel, Knochen und des lokalisierten Blutkreislaufsystems. Ungefähr drei von fünf Beschäftigten in der

Europäischen Union haben solche Beschwerden. Die am häufigsten auftretenden Arten sind dabei Rückenschmerzen und Muskelschmerzen in den oberen Gliedmaßen. Schlechte Haltung, Arbeiten in unangenehmen Positionen, schwere körperliche Arbeit, Heben von Gegenständen, sich wiederholende Tätigkeiten, Vibrationen durch Handwerkzeuge und niedrige Temperaturen lösen oft arbeitsbedingte Erkrankungen des Bewegungsapparates aus [15].

2.1.2 Methoden zur Erfassung der Ergonomie

Da sowohl für die verletzte Arbeitskraft als auch ihrem Unternehmen ein enormer Verlust entstehen kann, ist es notwendig, Risiken für Erkrankungen des Bewegungsapparates frühzeitig zu identifizieren und zu bewerten. Es wurden bereits einige Methoden entwickelt, um die Risikofaktoren zu ermitteln. Diese können in Selbstberichtsmethoden, Beobachtungsmethoden, direkte Messungen und indirekte Messungen eingeteilt werden. Biomechanische Modelle können diese Verfahren ergänzen, um die innere Beanspruchung in Muskeln und Gelenke zu ermitteln. In der Industrie werden derzeit typischerweise Selbstberichts- und Beobachtungsmethoden verwendet, allerdings basieren die Ergebnisse dieser Verfahren auf subjektiven Einschätzungen von Menschen und können sich somit von Person zu Person unterscheiden. Durch die Fortschritte bei der Sensortechnik entstanden neue Möglichkeiten, eine objektive Einschätzung der Ergonomie zu realisieren. Beispielsweise ermöglichen Goniometer, Kraftsensoren, Beschleunigungsmesser und Elektromyographie durch das Anbringen der Sensoren am Menschen eine direkte Messung der Körperhaltung. Mit Kameras können indirekte Messungen durchgeführt werden [1].

2.1.2.1 Selbstbericht

Selbstberichte umfassen Methoden wie Arbeitstagebücher, Interviews und Fragebögen. Diese Verfahren sind unkompliziert zu verwenden, auf eine breite Palette von Arbeitssituationen anwendbar und für die Befragung einer großen Anzahl an Personen bei vergleichsweise geringen Kosten geeignet. Ein weiterer Vorteil der Selbstberichterstattung besteht darin, dass die Beschäftigten Probleme melden können, die von anderen nur schwer beobachtet werden können, wie z.B. Schmerzen und die wahrgenommene Arbeitsbelastung. Das Hauptproblem ist jedoch, dass die Ergebnisse auf subjektiven Bewertungen beruhen und daher zwischen den einzelnen Personen erheblich variieren können. Darüber hinaus können die Antworten in den Umfragen aufgrund persönlicher Implikationen, Verständnisproblemen oder Interpretationsspielräumen verzerrt sein, was die Zuverlässigkeit dieser Methode einschränkt. Dieser Effekt kann durch eine Vergrößerung der Stichprobe gemindert werden, um sicherzustellen, dass die gesammelten Daten repräsentativ für die untersuchten Arbeitsplätze sind. Dadurch können aber nachfolgend hohe

Analysekosten entstehen und es sind geeignete Fähigkeiten notwendig, um die Ergebnisse genau zu interpretieren. Diese Methode eignet sich für eine erste Einschätzung, in der Arbeitsplätze mit vergleichsweise höherem Risiko identifiziert und dann einer detaillierteren Analyse unterzogen werden können [1] [16].

2.1.2.2 Beobachtung

Bei der Beobachtung werden die Körperhaltungen der Arbeitskräfte hinsichtlich Region, Häufigkeit, Schweregrad und Dauer durch eine erfahrene beobachtende Person systematisch aufgezeichnet. Dafür existieren einige Bewertungsformulare und -instrumente, mithilfe derer die Risikofaktoren identifiziert und beurteilt werden können. Derzeit stützt die Analyse der Ergonomie oft auf den Beobachtungen der Fachleute. Sie rufen nur eine minimale Störung der Arbeitsaufgabe hervor und ermöglichen somit eine Bewertung der Tätigkeiten unter realen Umgebungsbedingungen. Des Weiteren sind nur wenige zusätzliche Instrumente notwendig, weshalb diese Methode geringe Kosten verursacht. Allerdings beruht dieses Verfahren wie der Selbstbericht auf einer subjektiven Einschätzung, die bei unterschiedlichen Fachleuten zu verschiedenen Ergebnissen führen können. Möglicherweise müssen auch externe Fachleute beauftragt werden, da die erforderlichen Kompetenzen in dem jeweiligen Unternehmen nicht vorhanden sind. Darüber hinaus ist keine langfristige Beobachtung möglich, weshalb selten auftretende Phänomene nicht berücksichtigt werden können. Es existiert bereits eine Reihe solcher Verfahren. Im Folgenden werden die OWAS- und die RULA-Methode näher vorgestellt [1].

Ein finnisches Stahlunternehmen entwickelte das Bewertungsverfahren Ovako Working Posture Analysing System (OWAS). Hierbei wird die Haltung der Arme, der Beine und des Rückens mit vorgegebenen Posen verglichen. Insgesamt ergeben sich 84 mögliche Kombinationen für die Körperhaltung (dargestellt in Abbildung 1). Jeder Haltung eines Körperteils ist eine Nummer zugeordnet. Somit kann die gesamte Körperhaltung durch einen dreistelligen Code dargestellt werden. Das Beispiel auf der rechten Seite in Abbildung 1 ist somit durch den Code 215 beschrieben. Zusätzlich existieren auch noch 3 Kategorien für die getragene Last der Arbeitskraft. Anschließend können die Ergebnisse in vier Klassen eingeteilt werden. Die Klassen reichen von „normalen Körperhaltungen, die keine besondere Aufmerksamkeit erfordern, außer in einigen besonderen Fällen“ bis „Körperhaltungen müssen sofort berücksichtigt werden“ [17].

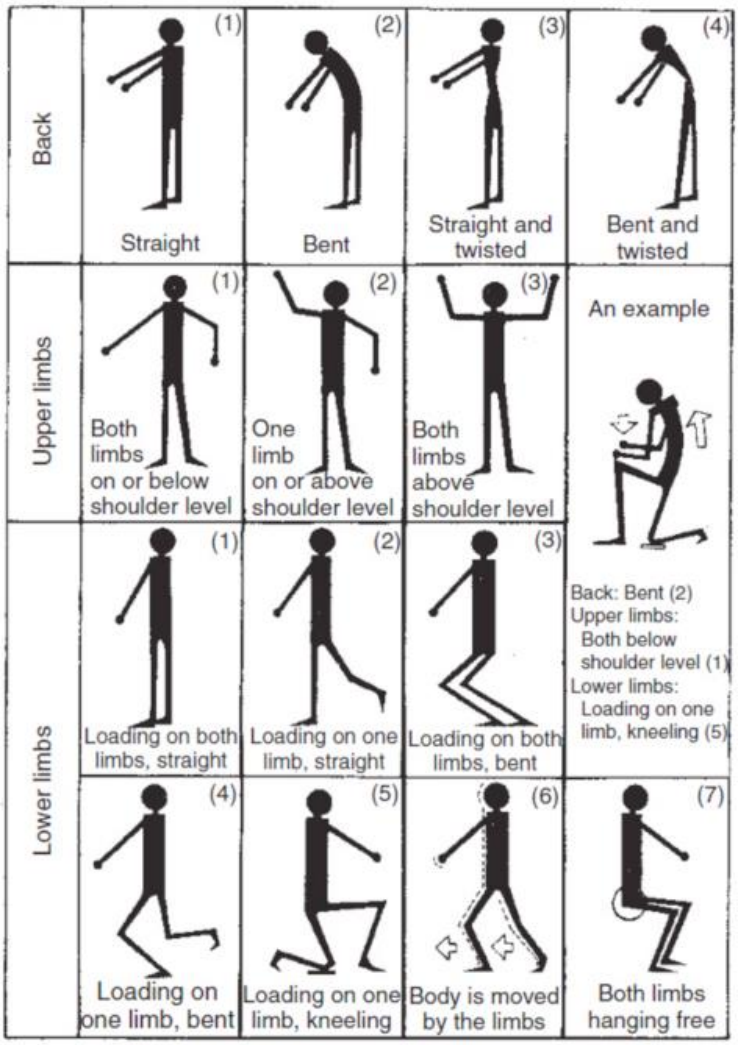

Abbildung 1 | Kategorien der Körperhaltungen bei der OWAS-Methode [17]

Das Rapid Upper Limb Assessment (RULA) ist ein weiteres Verfahren zur Beurteilung der Ergonomie am Arbeitsplatz mit speziellem Fokus auf die oberen Extremitäten. Diese Methode betrachtet die Arm- und Handgelenkhaltung sowie die Haltung von Oberkörper, Hals und Beine, wobei der Schwerpunkt auf der Arm- und Handgelenkhaltung liegt. Durch die genauere Betrachtung der Arme und Handgelenke und die Unterscheidung zwischen linkem und rechtem Arm ergeben sich bei der RULA-Methode 360 mögliche Haltungskombinationen. In die Bewertung fließt zusätzlich noch ein, ob Muskelarbeit, eine hohe Anzahl an Wiederholungen oder die Handhabung von Lasten notwendig ist. Mittels mehrerer Tabellen kann so ein Gesamtscore gebildet werden, welcher Werte von eins bis sieben umfasst. Basierend auf diesem Gesamtscore gibt die RULA-Methode Empfehlungen für weiteres Vorgehen. Arbeitsblätter, wie in Abbildung 2 dargestellt, erleichtern die Durchführung der Bewertung [8].

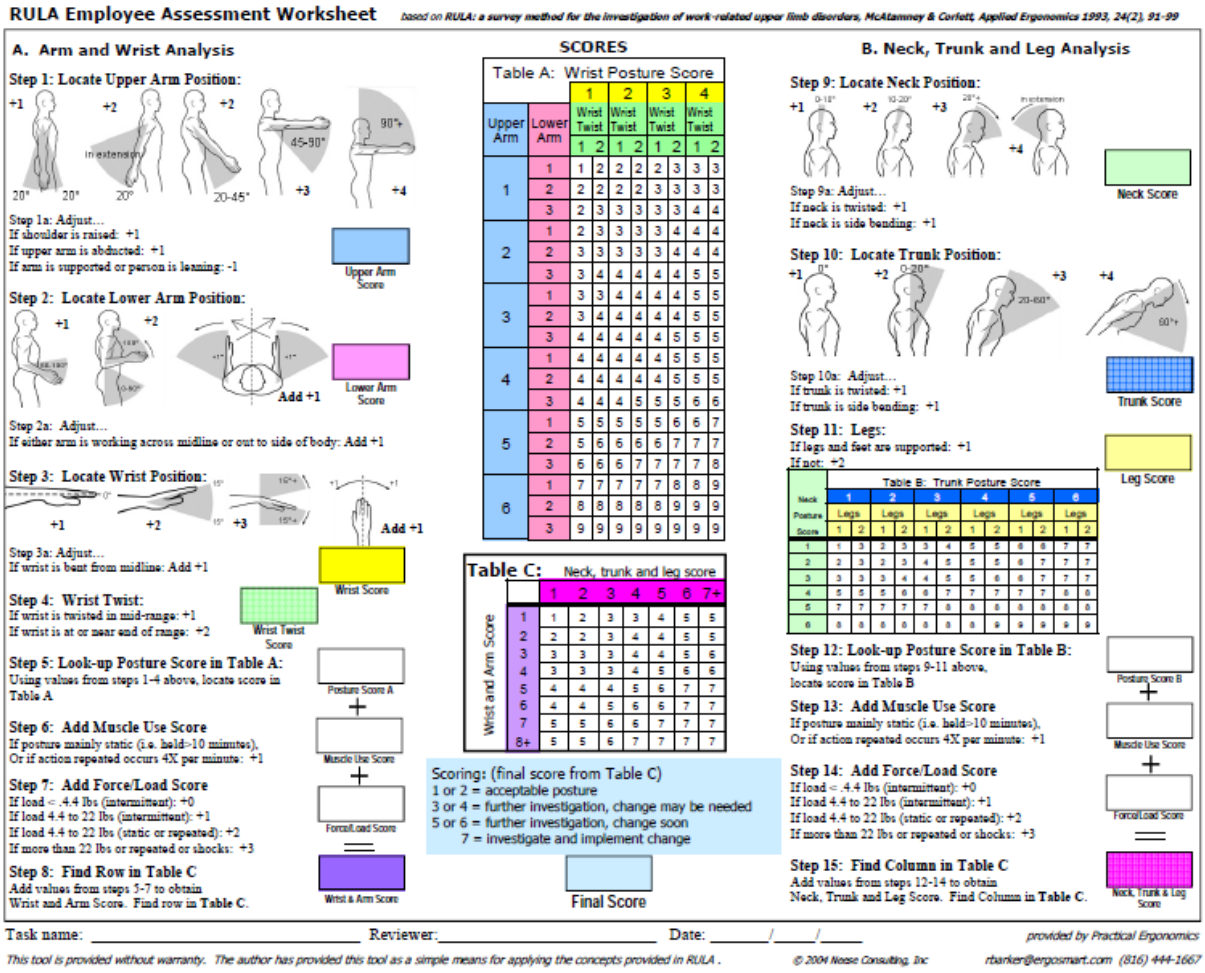

Abbildung 2 | Arbeitsblatt zur Beurteilung der Ergonomie mit der RULA-Methode [18]

2.1.2.3 Direkte Messung

Es gibt auch einige Methoden, um die Bewegungen des Körpers direkt zu messen. Dafür werden Marker oder Sensoren entweder direkt auf der Haut oder auf der Kleidung des Menschen aufgebracht, um die Bewegung von Gelenken und Körpersegmenten aufzuzeichnen. Als Sensoren kommen beispielsweise Goniometer (zur Messung von Winkeln), Kraftsensoren, Neigungs- und Beschleunigungsmesser zum Einsatz. Auch eine Messung der elektrischen Muskelaktivität mittels Elektromyographie ist möglich [1].

Eine sogenannte Inertial Measurement Unit (IMU) ist ein System aus mehreren Sensoren und kombiniert die Daten von Beschleunigung, Winkelgeschwindigkeit und magnetischem Feld zur Bestimmung der relativen Position eines Körperteils. Hierbei betrachtet man den menschlichen Körper als ein Modell, welches aus mehreren starren Segmenten besteht, die durch Gelenke verbunden sind. Die IMU kann dann auf jedem Körperteil platziert werden. Durch eine gleichzeitige Überwachung aller IMUs lässt sich die Haltung des Menschen rekonstruieren und somit können

Rückschlüsse auf die Ergonomie gezogen werden. Abbildung 3 zeigt eine beispielhafte Anordnung der IMUs am menschlichen Körper [19].

Der Vorteil dieser tragbaren Messeinheiten ist, dass sie, im Gegensatz zur Analyse mit kamerabasierten Systemen, unabhängig von Beleuchtungsänderungen und Verdeckungen durch andere Objekte sind [2].

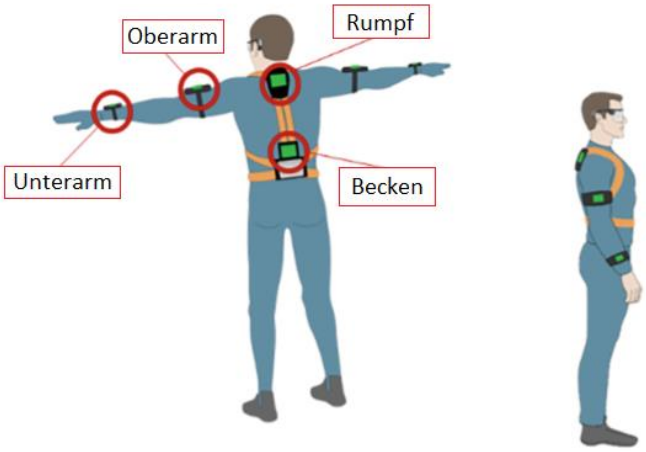

Abbildung 3 | Anordnung der IMUs für die Erfassung des Oberkörpers [19]

Eine andere Möglichkeit wäre, aktive oder passive Marker an bestimmten Teilen des menschlichen Körpers anzubringen. Bei passiven Markern handelt es sich meist um kleine, reflektierende Kugeln. Um diese Marker zu erkennen, benötigen die Kameras eigene (Infrarot-)Lichtquellen. Aktive Marker (wie in Abbildung 4) strahlen selbst ein Signal aus. Daher brauchen sie allerdings auch eine eigene Stromversorgung. Mittels eines Bildverarbeitungsalgorithmus können die Marker aus den Kamerabildern herausgefiltert und anschließend die aktuelle Pose berechnet werden [5].

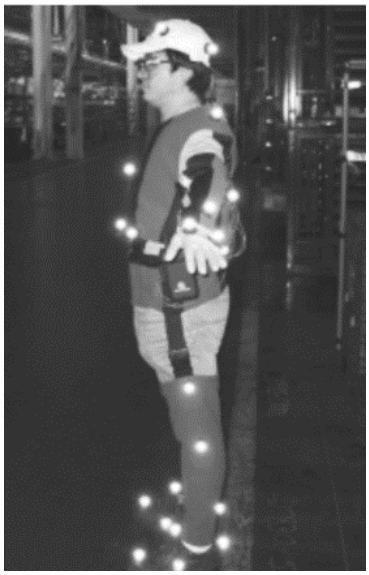

Abbildung 4 | Motion Capture-Anzug mit Markierungen an den Hauptgelenken [4]

Die direkte Messung wird häufig verwendet, um die Beobachtung durch Fachleute zu unterstützen oder zu ersetzen. Sie erreicht wesentlich höhere Genauigkeiten und ist, im Gegensatz zu Selbstbericht und Beobachtung, objektiv. Jedoch sind die Anfangsinvestitionen für die Geräte und der Aufwand für die Datenanalyse erheblich größer. Dadurch kann diese Methode bei einer großen Anzahl an Personen oder bei Erhebungen über eine längere Dauer nicht eingesetzt werden. Weiters sind für die Messungen oft Laborbedingungen notwendig, um störende Einflüsse zu minimieren. Das Verhalten der Arbeitskräfte kann aber auch durch die am Körper angebrachten Sensoren bzw. Marker beeinträchtigt werden und somit das Ergebnis verfälschen. Die direkte Messung eignet sich somit eher für die Beurteilung riskanter Körperbewegungen im Labor als für eine kontinuierliche Bewertung und Überwachung der Risiken direkt am Arbeitsplatz [1].

2.1.2.4 Indirekte Messung

Bei der indirekten Messung wird der Mensch auf Bildern bzw. Videos ohne die Verwendung von Markern oder Sensoren am Körper erkannt. Somit entfällt das Tragen eines Anzuges und die Bewegungen können unter realen Umgebungsbedingungen durchgeführt werden. Die indirekte Messung ruft dadurch nur eine geringe Störung bei der Arbeit hervor, ähnlich zu der Beobachtung durch einen Menschen. Im Gegensatz zur Beobachtung liefert sie aber objektive Resultate. Durch die Verwendung von Kameras ergeben sich allerdings auch einige Herausforderungen, wie z.B. Beleuchtungsänderungen oder Hindernisse in der Erfassungsrichtung [2].

Anfänglich erfolgte die Identifizierung der Gelenke zum Erhalten der kinematischen Daten wie Gelenkwinkel oder Beschleunigungen auf jedem Bild manuell [1]. Mittlerweile gibt es aber auch computergestützte Methoden. Sie basieren auf Computer-Vision-Algorithmen, welche den menschlichen Körper von der Hintergrundszene unterscheiden und die Gelenke erkennen [11]. Abbildung 5 zeigt das Ergebnis der Analyse eines zweidimensionalen Bilds mittels eines solchen Computer-Vision-Algorithmus. Aus einer zweidimensionalen Pose kann dann eine dreidimensionale Körperhaltung berechnet werden, aus welcher sich die Daten für Gelenkwinkel, Geschwindigkeiten und Beschleunigungen ermitteln lassen [20]. Die dreidimensionalen Informationen für die Körperhaltung können aber genauso von einer Tiefenkamera bezogen werden. Des Öfteren kommt auch ein Netzwerk von Kameras zum Einsatz, um die genaue Position der Arbeitskraft im Raum zu bestimmen [11].

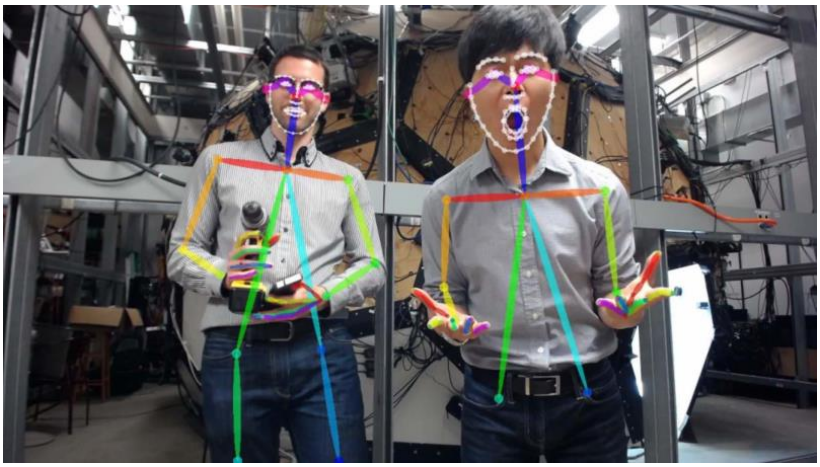

Abbildung 5 | Erkennung des Menschen aus einem zweidimensionalen Foto mittels künstlicher Intelligenz [12]

2.1.2.5 Biomechanische Modelle

Zu einer vollständigen Beurteilung der Risiken müssen eventuell auch die inneren Belastungen der Gelenke und Muskeln ermittelt werden, die durch Kräfte und Momente bei einer Bewegung entstehen. Das Ziel von biomechanischen Modellen ist, diese Belastung genau abzuschätzen. In den Modellen sind auch Grenzwerte definiert, welchen die Gelenke und das Gewebe standhalten können. Abbildung 6 zeigt ein Beispiel eines biomechanischen Modells. Im Gegensatz zu geometrischen Modellen, welche nur Eigenschaften wie Geschwindigkeit und Position beinhalten, integrieren biomechanische Modelle auch die muskulären Eigenschaften eines Körpers. Da sich die Menschen hinsichtlich Geschlecht, Alter und Gewicht unterscheiden, sind aber große Mengen an Daten erforderlich. Darüber hinaus besteht die Möglichkeit, dass die Gelenkpositionen vom biomechanischen Modell von den Positionen des Bewegungserfassungssystems abweichen, wodurch Fehler auftreten können [1].

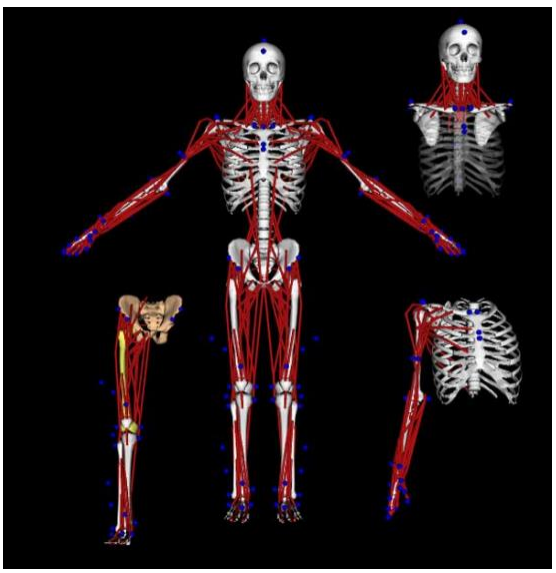

Abbildung 6 | Biomechanische Modelle in OpenSim (rote Linien sind Muskeln, blaue Punkte sind virtuelle Marker) [21]

Biomechanische Modelle werden heutzutage oft für die Nachbearbeitung der Bewegungsdaten verwendet, welche aus der direkten bzw. indirekten Messung folgen. Beispielsweise können die internen Belastungen durch Berechnungen mit biomechanischen Ganzkörpermodellen aus den - mit passiven Markern aufgenommenen - Geschwindigkeiten und Positionen der Gelenke ermittelt werden. Es existieren auch biomechanische Modelle, welche die Elektromyographie unterstützen. Diese benötigen Sensoren an den vorgesehenen Muskelstellen, um die jeweilige Muskelkraft zu messen. Eine von der direkten bzw. indirekten Messung unabhängige Nutzung für die Bewegungsanalyse des Menschen ist natürlich genauso möglich [1].

2.2 Künstliche neuronale Netzwerke

Dieses Kapitel beschreibt die Grundlagen und die Funktionsweise von künstlichen neuronalen Netzwerken. Anschließend wird noch näher auf den Spezialfall der faltenden neuronalen Netzwerke eingegangen. Diese kommen hauptsächlich in der Bildverarbeitung zum Einsatz und werden somit auch für die Schätzung der menschlichen Pose verwendet.

2.2.1 Grundlagen

Warren McCulloch und Walter Pitts definierten im Jahr 1943 erstmals künstliche neuronale Netze [22]. Dabei handelt es sich um Algorithmen, die dem menschlichen Gehirn nachempfunden sind. Dieses Modell ermöglicht es heutzutage, komplexe Aufgaben in vielen Bereichen zu lösen. Beispielsweise lassen sich dadurch verschiedene Datenquellen wie Bilder, Geräusche, Texte, Tabellen oder Zeitreihen interpretieren, um Informationen oder Muster zu erkennen. Diese können auf unbekannte Daten angewendet werden, um Vorhersagen für die Zukunft zu erstellen. Künstliche neuronale Netzwerke finden dort Anwendung, wo wenig systematisches Wissen vorliegt, aber eine große Menge unpräziser Eingabeinformationen (unstrukturierte Daten) verarbeitet werden müssen, um ein konkretes Ergebnis zu erhalten. Dies ist der Fall beim autonomen Fahren, bei der Bilderkennung, beim Übersetzen von Sprachen, bei Wettervorhersagen, bei Krankheitsanalysen und vielem mehr [23].

Künstliche neuronale Netzwerke weisen im Wesentlichen die Strukturen gerichteter Graphen auf. Sie bestehen aus vielen Knoten (auch Neuronen genannt), welche mittels gewichteten Verbindungen ein Netz bilden. Die Knoten sind Eingabe-, Ausgabe- und versteckten Schichten zugeordnet. In der Eingabeschicht werden die zu verarbeitenden externen Reize und Variablen aufgenommen. Die Ausgabeschicht stellt die Ergebnisse dar. Die versteckten Knoten dazwischen empfangen, verarbeiten und leiten die Signale weiter. Während die Ein- und die Ausgabeschicht lediglich aus

einer Ebene bestehen, können die versteckten Knoten in beliebig vielen Ebenen angeordnet werden. Der Prozess im Inneren ist nicht sichtbar und gleicht somit einer Black Box, weshalb diese Ebenen auch versteckte Schichten genannt werden. Eine Übersicht der Schichten bietet Abbildung 7 [23] [24].

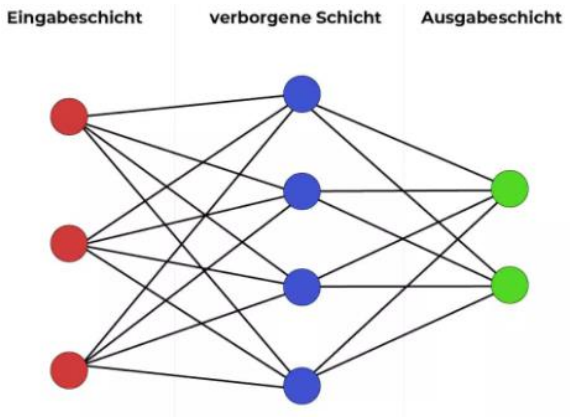

Abbildung 7 | Schichten eines neuronalen Netzwerks [23]

Es existieren viele verschiedene Typen von neuronalen Netzwerken. Sie können keine oder mehrere versteckte Schichten beinhalten. Liegen besonders viele versteckte Schichten vor, spricht man von Deep Learning. Bei vorwärtsgerichteten Netzwerken (Feedforward Neural Networks, FNN) sind die Knoten einer Schicht lediglich mit allen Knoten der nächsten Schicht verbunden. Rekurrente neuronale Netze (Recurrent Neural Networks, RNN) besitzen im Gegensatz dazu auch rückwärtsgerichtete Verbindungen. Oft werden diese Rückkopplungen mit einer Zeitverzögerung versehen, sodass ein Gedächtnis entsteht. In Abbildung 8 sind die grundlegenden Netzstrukturen dargestellt [23] [25].

Eine weitere wichtige Gruppe sind die faltenden neuronalen Netzwerke (Convolutional Neural Networks, CNN). Sie arbeiten besonders effizient mit zwei- und dreidimensionalen Eingabedaten und werden beispielsweise für die Objekterkennung in Bildern und Videos eingesetzt. Da diese Diplomarbeit auf solchen faltenden neuralen Netzen basiert, wird im nächsten Kapitel genauer darauf eingegangen [23].

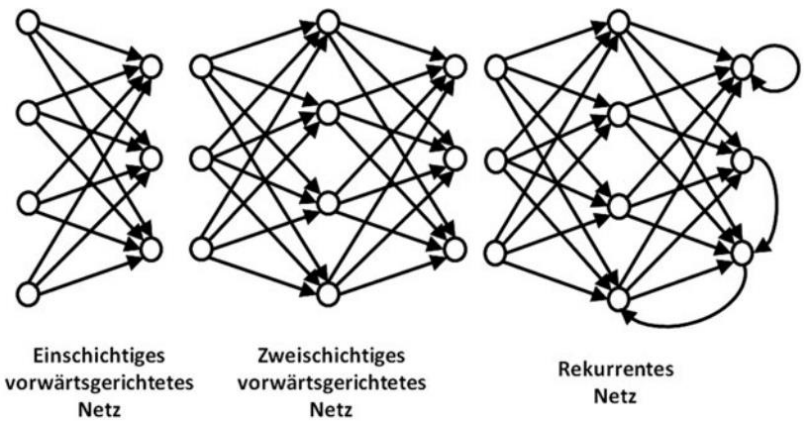

Abbildung 8 | Architekturprinzipien von neuronalen Netzen [25]

Der Wert eines Neurons ergibt sich dabei aus der Summe der gewichteten Eingangssignalen und der sogenannten Aktivierungsfunktion. Die Aktivierungsfunktion dient dazu, das Ergebnis auf einen bestimmten kontinuierlichen Wertebereich (z.B. -1 bis 1) abzubilden. Das Neuron gibt bei Überschreiten eines bestimmten Schwellenwerts das Signal dann weiter. Abbildung 9 zeigt eine Übersicht über die Vorgänge und Elemente eines künstlichen Neurons [24].

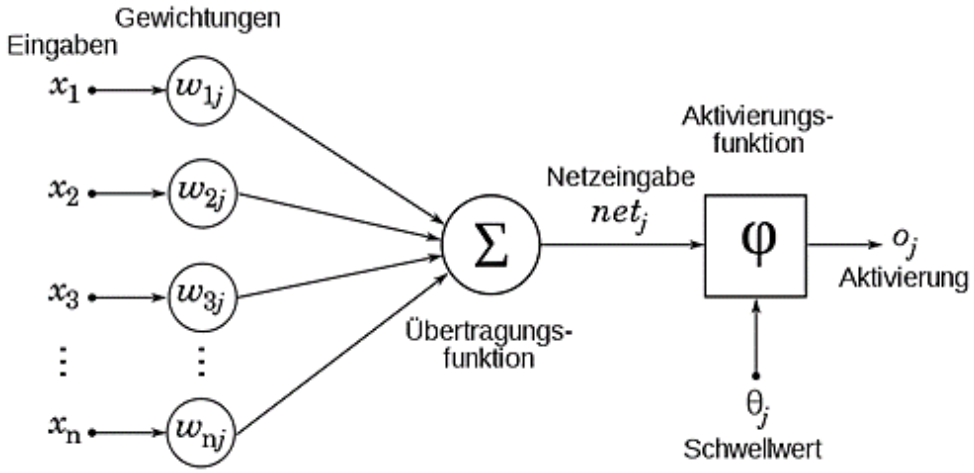

Abbildung 9 | Darstellung der Elemente eines künstlichen Neurons [23]

Das Training der künstlichen neuronalen Netzwerke erfolgt meist durch Anpassung der Gewichte. Das Ziel dieses Lernvorganges ist, den Fehler zwischen erwarteter Ausgabe und tatsächlicher Ausgabe sukzessive zu minimieren. Anfangs sind den Gewichten zufällige Werte zugeteilt. Das System beinhaltet noch kein bestimmtes Wissen. Beim Lernvorgang werden dem Netz verschiedene bekannte Daten angeboten. Nach jedem Durchlauf passt ein Korrektursignal die Gewichte so an, dass der gesamte Fehler des Netzwerks Schritt für Schritt minimiert wird. Um ein zufriedenstellendes Ergebnis zu erhalten, muss der Trainingsdatensatz ausreichend groß sein [24].

Ein wesentlicher Vorteil von künstlichen neuronalen Netzwerken besteht darin, dass dem System keine Vorgaben im Sinne vorgeformter Regeln gemacht werden müssen. Das Netzwerk extrahiert selbstständig die Informationen aus den gegebenen Lernbeispielen und generiert auch selbstständig eine Lösung. Dabei können auch sehr gut nichtlineare und komplexe Zusammenhänge realisiert werden. Gerade dies stellt einen großen Vorteil gegenüber den klassischen mathematischen Methoden dar. Die Regeln, die das künstliche neurale Netzwerk aus den Datensätzen generiert hat, sind allerdings schwer ableitbar, weshalb die Ergebnisse nur bedingt nachvollzogen werden können. Der relative Einfluss eines Eingabefaktors auf das Endergebnis lässt sich somit nicht genau bestimmen. Des Weiteren ist es durch die große Anzahl an Einstellungsmöglichkeiten denkbar, dass die gewählten Netzwerkeinstellungen nicht optimal sind und noch bessere Einstellungen existieren. Tabelle 1 zeigt eine Übersicht der Vor- und Nachteile von künstlichen neuronalen Netzwerken [24].

Vorteile	Nachteile
Lernfähigkeit und Generalisierungsfähigkeit anhand von Beispielen und Mustern	Probleme im Training ("overfitting")
Kein Vorwissen über bestehende Zusammenhänge erforderlich	Bei komplexen Problemen unter Umständen sehr lange Rechendauer (bis zu Tagen)
Gute Modulierung nichtlinearer Zusammenhänge	Keine mathematische Präzision bei Ergebnisprognosen
Gute Modellierung beliebiger Interaktionen zweier Variablen	Kausale Beziehungen und logische Schlussfolgerungen nur sehr bedingt formulierbar
Anpassungsfähigkeit (bei Änderungen des beobachteten Prozesses können jederzeit neue Daten präsentiert werden)	Daten der Test- und der Anwendungsphase müssen sich ähneln
Robustheit (verzerrte oder unvollständige Daten können mit akzeptabler Genauigkeit verarbeitet werden)	Ergebnisse eines künstlichen neuralen Netzwerks haben "Black-Box-Charakter", sodass die Ergebnisse von Anwendenden eventuell nur bedingt akzeptiert werden

Tabelle 1 | Vor- und Nachteile eines künstlichen neuronalen Netzwerkes [24]

2.2.2 Faltende neuronale Netzwerke

Anwendung finden faltende neuronale Netze (Convolutional Neural Networks, CNN) im Bereich der Klassifizierung und Musteranalyse, wie z.B. bei der Gesichts-, Bild- oder Spracherkennung. Fahrerassistenzsysteme nutzen faltende neuronale Netzwerke beispielsweise, um die Fahrspur und Verkehrszeichen zu erkennen. Bei Google werden sie verwendet, um in Google StreetView die Nummernschilder und Gesichter unkenntlich zu machen [25].

Faltende neuronale Netzwerke sind vorwärtsgerichtete Netze, allerdings sind hier die benachbarten Schichten nicht vollständig miteinander verbunden. Jedes Neuron bekommt als Eingang nur einen kleinen zusammenhängenden Bereich der vorherigen Schicht. Durch diese deutlich geringere Anzahl an Verbindungen, sind bei der Anwendung erheblich weniger Rechenschritte notwendig und der Lernvorgang erfolgt schneller. Dies ist insofern wichtig, da Bilder aus einer Vielzahl an Pixeln bestehen und in der Eingabeschicht jedem Pixelwert ein Neuron zugewiesen wird. Damit würden bei der Verarbeitung von Bildern mit klassischen FNNs große Mengen an Gewichtungen zusammenkommen, die durch die lokale Verknüpfung besser bewältigt werden können. Mathematisch betrachtet, entspricht das Prinzip einer diskreten Faltung, woher sich auch der Name ableiten lässt [25].

Das faltende neuronale Netzwerk besteht aus sogenannten Faltungsschichten (Convolutional Layer) und Poolingschichten, welche nach der Eingabeschicht angeordnet sind. Am Ende befinden sich eine oder mehrere vollständig verbundene Schichten (Fully Connected Layer). Abbildung 10 gibt eine Übersicht über den Aufbau von faltenden neuronalen Netzwerken [25].

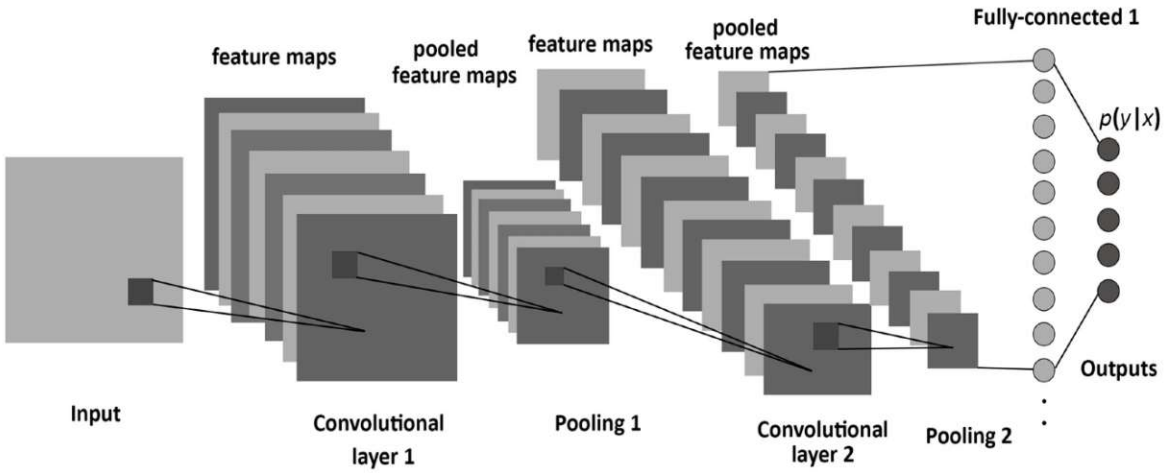

Abbildung 10 | Struktur von faltenden neuronalen Netzwerken, bestehend aus Faltungs-, Pooling- und vollständig verbundenen Schichten [26]

Anfangs entspricht die Anzahl der Pixel multipliziert mit den Farbkanälen des Bilds der Anzahl der Neuronen. Nachdem der Eingabeschicht ein Bild übergeben und jedem Neuron die Farbintensität zugeordnet wurde, wird die Eingabe mit einem Faltungskern (Filter) gefaltet. Dieser Faltungskern ist kleiner als die vorherige Schicht und wird wie ein Fenster über das gesamte Bild gezogen. Beim Lernen des Netzwerks werden die Werte der Faltungskerne angepasst [25].

Abbildung 11 beschreibt links den Vorgang des Faltens anhand eines Beispiels. Durch einen Faltungskern mit dem Format 3x3 ergibt sich aus dem Eingabebild der Dimension 4x4 ein Ausgabebild vom Format 2x2. Der erste Ausgabewert berechnet sich durch elementweise Multiplikation des Faltungskerns mit der linken oberen 3x3 Matrix. Der nächste Wert ergibt sich durch Verschieben des Faltungskerns um eine Position. Liegen mehrere Eingangsbilder vor, erhält man das Ausgabebild durch die Summe der gefalteten Bilder (siehe Abbildung 11 rechts). Für drei Eingabe- und zwei Ausgabebilder werden 6 Faltungskerne benötigt [25].

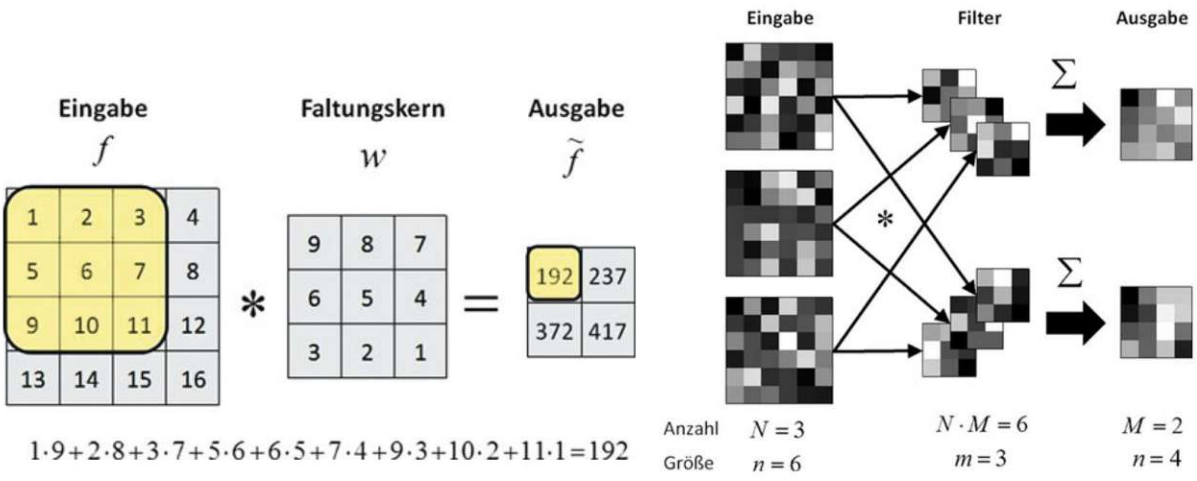

Abbildung 11 | Faltung eines Bildes [25]

In den Faltungskernen sind verschiedene Merkmale, wie Linien oder bestimmte Formen dargestellt. Durch die Faltung wird überprüft, ob die jeweilige Form in dem Eingabebild vorhanden ist. Während die erste Ebene nur einfache Formen wie Linien oder Kurven beinhaltet, können in den darauffolgenden Faltungsschichten immer komplexere Konturen erfasst werden. Dadurch ist die Faltungsschicht in der Lage, einzelne Merkmale zu erkennen und zu extrahieren. Die Ausgabebilder werden deshalb auch als Merkmalskarten (Feature Maps) bezeichnet. Abbildung 12 zeigt beispielhaft die Merkmalskarten, die aus drei verschiedenen Faltungskernen resultieren [27].

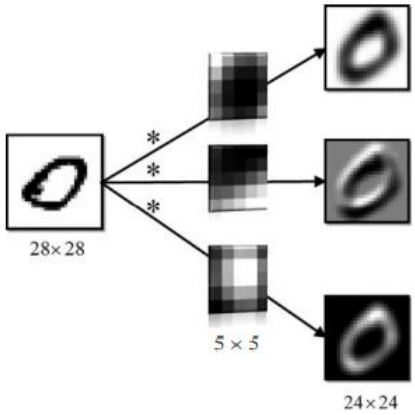

Abbildung 12 | Beispiele für Merkmalskarten (Feature Maps) [25]

Anschließend werden die gefalteten Bilddaten in der Poolingschicht mithilfe spezieller Pooling-Operatoren (z.B. Maximal-Pooling oder Mittelwert-Pooling) in der Dimension reduziert. Abbildung 13 zeigt das Maximal-Pooling anhand eines Beispiels. Bei Pooling-Faktor 2 wird das Bild in sich nicht überschneidende Bereiche der Dimension 2x2 aufgeteilt. Beim Maximal-Pooling wird aus diesem Bereich dann der betragsmäßig größte Wert entnommen und mit der zugehörigen Position gespeichert. Die Poolingschicht verdichtet und reduziert also die Auflösung der Ergebnisse und dient dazu, nur die relevantesten Signale an die nächsten Schichten weiter zu geben. Durch die reduzierte Anzahl an Daten erhöht sich des Weiteren die Berechnungsgeschwindigkeit [25] [27].

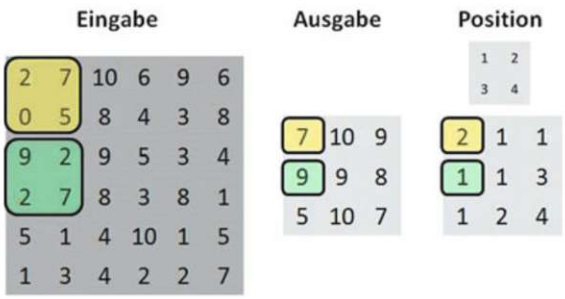

Abbildung 13 | Maximal-Pooling mit Pooling-Faktor 2 [25]

Nach der letzten Poolingschicht befinden sich noch eine oder mehrere vollständig verbundene Schichten, bei denen alle Neuronen direkt mit allen Neuronen der jeweils benachbarten Schicht verbunden sind. Dies entspricht den versteckten Schichten in traditionellen künstlichen neuronalen Netzwerken. Die vollständig verbundenen Schichten am Ende dienen hauptsächlich zur Klassifizierung. Die Anzahl der Neuronen in der letzten Schicht korrespondiert dann üblicherweise zu der Anzahl der Objektklassen, die das faltende neurale Netzwerk unterscheiden soll. Die Ausgabeschicht gibt anschließend die Ergebnisse aus [25] [27].

2.3 Schätzung der menschlichen Pose

Bei der Schätzung der menschlichen Pose (Human Pose Estimation, HPE) handelt es sich um die Erkennung der Anordnung menschlicher Körperteile auf Eingabedaten wie Bildern und Videos. Sie liefert Informationen bezüglich der Geometrie und der Bewegung des menschlichen Körpers, die vielseitig angewendet werden können (z.B. Mensch-Maschine-Interaktion, Bewegungsanalyse, Augmented Reality, Virtual Reality, Gesundheitswesen usw.). Durch den Einsatz von Deep-Learning-Techniken, insbesondere den faltenden neuronalen Netzwerken, wurden in den letzten Jahren in diesem Bereich erhebliche Fortschritte erzielt. Jedoch führen Überschneidungen, unzureichende Trainingsdaten und Tiefenmehrdeutigkeit immer noch oft zu Problemen [28].

Grundsätzlich wird aufgrund der unterschiedlichen Herausforderungen zwischen der 2D- und 3D- sowie der Einzelpersonen- und der Mehrpersonen-Posenschätzung unterschieden. Bei der 2D-Posenschätzung werden für die Pose einer einzelnen Person durch die Verwendung von Deep-Learning-Techniken bereits gute Ergebnisse erzielt. Im Gegensatz dazu ist es bei der 3D-Posenschätzung viel schwieriger, genaue 3D-Daten zu erhalten. Hier besteht die größte Herausforderung in der Mehrdeutigkeit der Tiefe. Oft werden Sensoren wie Tiefensensoren, Inertial Measurement Units (IMUs) und Hochfrequenzgeräte verwendet, aber diese Ansätze sind normalerweise nicht kosteneffektiv und erfordern spezielle Hardware [28].

Um die Ergebnisse darzustellen und die Haltung des menschlichen Körpers zu beschreiben, wird meist ein starres kinematisches Modell (Skelettmodell) verwendet, bei dem die Gelenke, auch Keypoints genannt, mittels Linien verbunden sind. Dieses flexible und intuitive menschliche Körpermodell kommt sowohl bei der 2D- als auch bei der 3D-Posenschätzung zur Anwendung. Jedoch ist es bei der Darstellung von Textur- und Forminformationen begrenzt, weshalb auch planare Modelle für die 2D- und volumetrische Modelle für die 3D-Posenschätzung existieren. Im planaren Modell werden Körperteile normalerweise durch Rechtecke dargestellt, die sich den Konturen des menschlichen Körpers annähern. Das volumetrische Modell bildet den Menschen

vollständig und realistisch ab. Abbildung 14 zeigt die drei verschiedenen Arten von Menschmodellen [28].

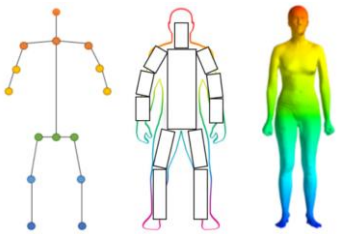

Abbildung 14 | Kinematisches, planares und volumetrisches Modell des Menschen [28]

2.3.1 2D - Posenschätzung

2.3.1.1 2D - Einzelpersonen - Posenschätzung

Die 2D-Posenschätzung kann wiederum in Einzelpersonen- und Mehrpersonen-Schätzung eingeteilt werden. Die 2D-Einzelpersonen-Posenschätzung wird verwendet, um Gelenkpositionen des menschlichen Körpers zu lokalisieren, wenn die Eingabe ein Bild mit einer einzigen Person ist. Sind mehrere Personen zu sehen, wird das Eingabebild zuerst zugeschnitten, sodass sich in jedem zugeschnittenen Unterbild nur eine Person befindet. Dieser Vorgang kann automatisch mittels eines Oberkörper- oder Ganzkörperdetektors durchgeführt werden. Im Allgemeinen unterscheidet man bei den Einzelpersonen-Methoden zwischen Regressionsmethoden und Körperteilerkennungsmethoden zur Erkennung von Körperteilen. Abbildung 15 zeigt den Ablauf der beiden Methoden zur 2D-Einzelpersonen-Posenschätzung [28].

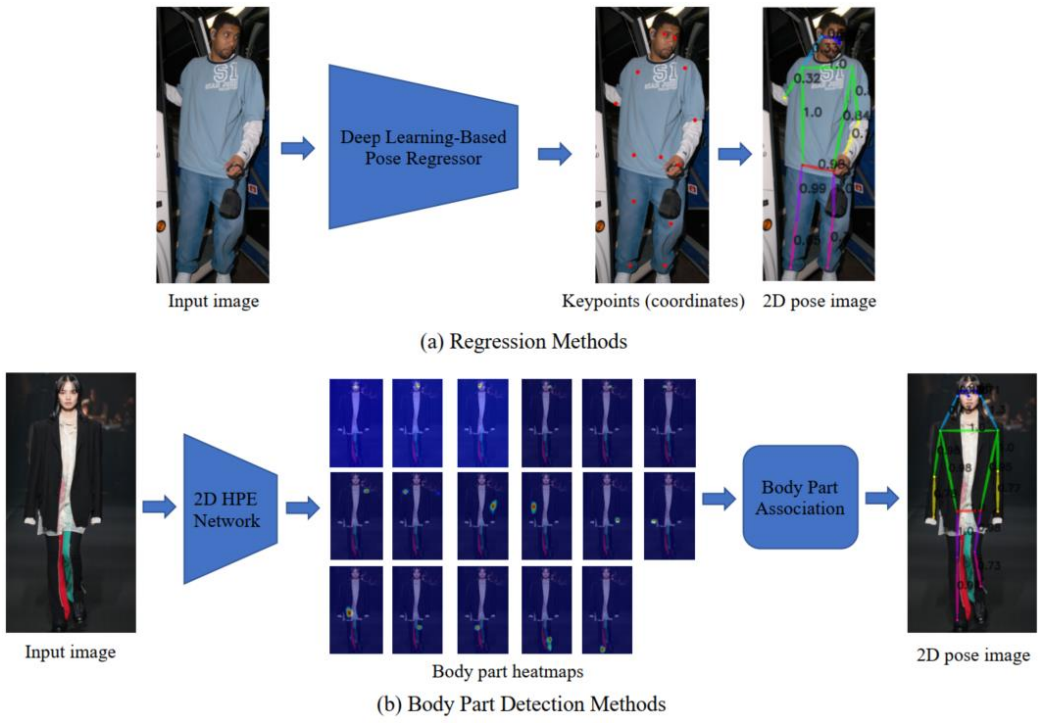

Abbildung 15 | Ablauf der Regressions- (a) und der Körperteilerkennungsmethode (b) [28]

Regressionsmethoden wenden ein tiefes neuronales Netzwerk an, um Eingabebilder auf das kinematische Körpermodell abzubilden und Keypoint-Koordinaten zu erzeugen. Dieses bietet ein schnelles Lernen und eine Vorhersagegenauigkeit auf Pixel-Ebene. Aufgrund des stark nichtlinearen Problems ergeben sich jedoch in der Regel suboptimale Lösungen [28].

Ziel der Körperteilerkennungsmethoden ist es, zuerst die wahrscheinliche Position von Körperteilen und Gelenken vorherzusagen und diese mittels Heatmaps darzustellen. Diese Daten bilden die Grundlage für die anschließende Posenschätzung. Insbesondere auf Heatmaps basierende Körperteilerkennungsmethoden werden in 2D-Posenschätzungen häufig verwendet, da die Vorhersage jedes Pixels in der Heatmap die Genauigkeit der Lokalisierung der Keypoints verbessern kann. Die Genauigkeit der vorhergesagten Keypoints hängt jedoch von der Auflösung der Heatmaps ab. Der Rechenaufwand und der Speicherbedarf werden bei Verwendung hochauflösender Heatmaps erheblich erhöht [28].

2.3.1.2 2D - Mehrpersonen - Posenschätzung

Im Vergleich zur Einzelpersonen-Posenschätzung ist die Mehrpersonen-Posenschätzung schwieriger und herausfordernder, da die Anzahl der Personen und ihre Positionen ermittelt sowie die Keypoints für jede Person gruppiert werden müssen. Die Methoden für mehrere Personen können in Top-Down- und Bottom-Up-Methoden unterteilt werden [28].

Top-Down-Verfahren erfassen zuerst mittels Erkennungsmethoden jede Person auf dem Bild und schätzen dann die Positionen der Keypoints mithilfe der auf einzelnen Personen basierenden Ansätze. Bei dieser Methode ist die Keypoint- Schätzung innerhalb jeder erfassten Personenregion wesentlich einfacher, weil der Hintergrund weitgehend entfernt wird. Dadurch liefert die Top-Down-Methode oft bessere Ergebnisse [28].

Im Gegensatz zu Top-Down-Methoden erkennen Bottom-Up-Methoden zuerst alle Keypoints in einem Bild und gruppieren sie dann zu einzelnen Posen. Die Rechengeschwindigkeit ist bei Bottom-Up-Methoden normalerweise schneller als bei Top-Down-Methoden, da sie die Pose nicht für jede Person separat ermitteln müssen. Im Top-Down-Ansatz wirkt sich die Anzahl der Personen im Eingabebild direkt auf die Rechenzeit aus. Abbildung 16 zeigt den Ablauf der beiden Methoden zur 2D-Mehrpersonen-Posenschätzung [28].

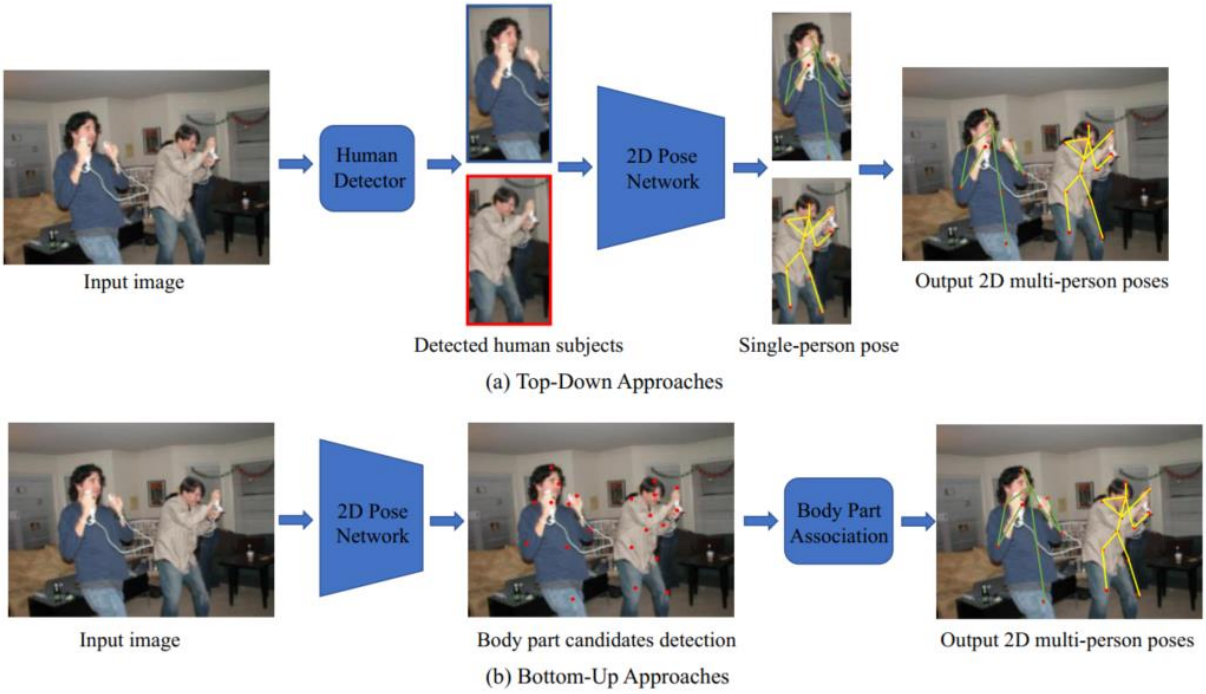

Abbildung 16 | Ablauf von Top-Down- (a) und Bottom-Up-Methoden (b) [28]

Bei der 2D-Posenschätzung für mehrere Personen existieren derzeit einige Herausforderungen. Bei großen Überschneidungen ist die zuverlässige Erkennung von Personen schwierig. Die Personendetektoren in Top-Down-Methoden können möglicherweise die Grenzen weitgehend überlappender menschlicher Körper nicht identifizieren. Auch für Bottom-Up-Ansätze ist die Keypoint-Gruppierung bei starken Überschneidungen schwierig. Eine weitere Herausforderung ist die Berechnungseffizienz. Obwohl einige Methoden eine Echtzeitverarbeitung erreichen können, ist es immer noch schwierig, die künstlichen neuronalen Netzwerke auf Geräten mit eingeschränkten Ressourcen zu implementieren. Reale Anwendungen (z.B. Online-Coaching, Gaming, Augmented Reality und Virtual Reality) erfordern effiziente Methoden, die den Nutzenden ein besseres Interaktionserlebnis bieten können. Des Weiteren liegen oft begrenzte Trainingsdaten für seltene Posen vor. Obwohl die aktuellen Datensätze für die normale Posenschätzung (z.B. Stehen, Gehen, Laufen) groß genug sind, haben sie nur begrenzte Daten für ungewöhnliche Posen, wie z.B. das Fallen. Das kann zu einer schlechten Erkennung solcher Posen führen [28].

2.3.2 3D - Posenschätzung

Mit der 3D-Human Pose Estimation können die Positionen von Körpergelenken im 3D-Raum vorhergesagt und somit umfangeiche 3D-Strukturinformationen zum menschlichen Körper geliefert werden. Obwohl die 2D-Posenschätzung bereits gute Ergebnisse liefert, ist die 3D-Posenschätzung zurzeit noch eine herausfordernde Aufgabe. Da bei aus einem einzigen Winkel aufgenommenen Bildern und Videos die

3D-Umgebung auf zwei Dimensionen projiziert wird, ist das Problem nicht mehr eindeutig lösbar. Eine Lösung besteht darin, die menschliche 3D-Pose aus mehreren Ansichten abzuschätzen, da der verdeckte Teil aus anderen Perspektiven sichtbar werden kann. Um die 3D-Pose aus mehreren Ansichten zu rekonstruieren, muss allerdings die Zuordnung der entsprechenden Position zwischen den verschiedenen Kameras gelöst werden. Auch andere Sensoren wie IMUs können zur Vereinfachung des Problems beitragen. Eine weitere Einschränkung besteht darin, dass, im Gegensatz zu den menschlichen 2D-Datensätzen, das Sammeln genauer 3D-Posen zeitaufwändig und eine manuelle Zuweisung nicht praktikabel ist. In weiterer Folge wird die 3D-Posenschätzung aus einer Perspektive genauer betrachtet [28].

Die Rekonstruktion von menschlichen 3D-Posen aus einer einzigen Perspektive von Bildern und Videos ist eine nicht triviale Aufgabe, die von Überschneidungen durch den Menschen selbst, Überschneidungen durch andere Objekte, Tiefenmehrdeutigkeiten und unzureichenden Trainingsdaten sehr beeinträchtigt wird. Es ist ein stark mehrdeutiges Problem, da verschiedene menschliche 3D-Posen auf eine ähnliche 2D-Pose projiziert werden können. Darüber hinaus können bei Methoden, die auf 2D-Keypoints aufbauen, geringfügige Lokalisierungsfehler der Keypoints zu großen Posenverzerrungen im 3D-Raum führen. Im Vergleich zum Einzelpersonenszenario ist der Mehrpersonenfall wesentlich komplizierter [28].

2.3.2.1 3D - Einzelpersonen - Posenschätzung

3D-Posenschätzungsansätze für eine Person können in modellfreie und modellbasierte Kategorien eingeteilt werden, je nachdem, ob sie ein menschliches Körpermodell zur Schätzung der menschlichen 3D-Pose verwenden oder nicht. Abbildung 17 zeigt den Ablauf der jeweiligen Verfahren [28].

Die modellfreien Methoden verwenden keine menschlichen Körpermodelle, um die 3D-Darstellung des Menschen zu rekonstruieren. Diese Methoden können wiederum in zwei Klassen unterteilt werden. Zum einen in direkte Schätzungsansätze, welche die menschliche 3D-Pose direkt aus 2D-Bildern ableiten und zum anderen in indirekte Methoden, welche zwischenzeitlich die 2D-Pose schätzen und anschließend erst die 3D-Pose ermitteln. Durch die mittlerweile gut funktionierenden 2D-Posendetektoren übertreffen die 2D-to-3D-Lifting-Verfahren mit Zwischenschritt im Allgemeinen direkte Schätzansätze [28].

Modellbasierte Methoden beinhalten parametrische Körpermodelle (wie das kinematische Modell und das Volumenmodell), um die Haltung und die Form des Menschen abzuschätzen. Das kinematische Modell ist eine durch Knochen und Gelenke verbundene Körperdarstellung mit kinematischen Einschränkungen. Dadurch können Kenntnisse wie z.B. Informationen zur Konnektivität der Skelettgelenke, Eigenschaften der Gelenkrotation und feste Knochenlängenverhältnisse für eine

plausiblere Posenschätzung genutzt werden. Volumetrische Modelle sind zusätzlich dazu in der Lage, qualitativ hochwertige menschliche Darstellungen und Forminformationen des menschlichen Körpers zu liefern [28].

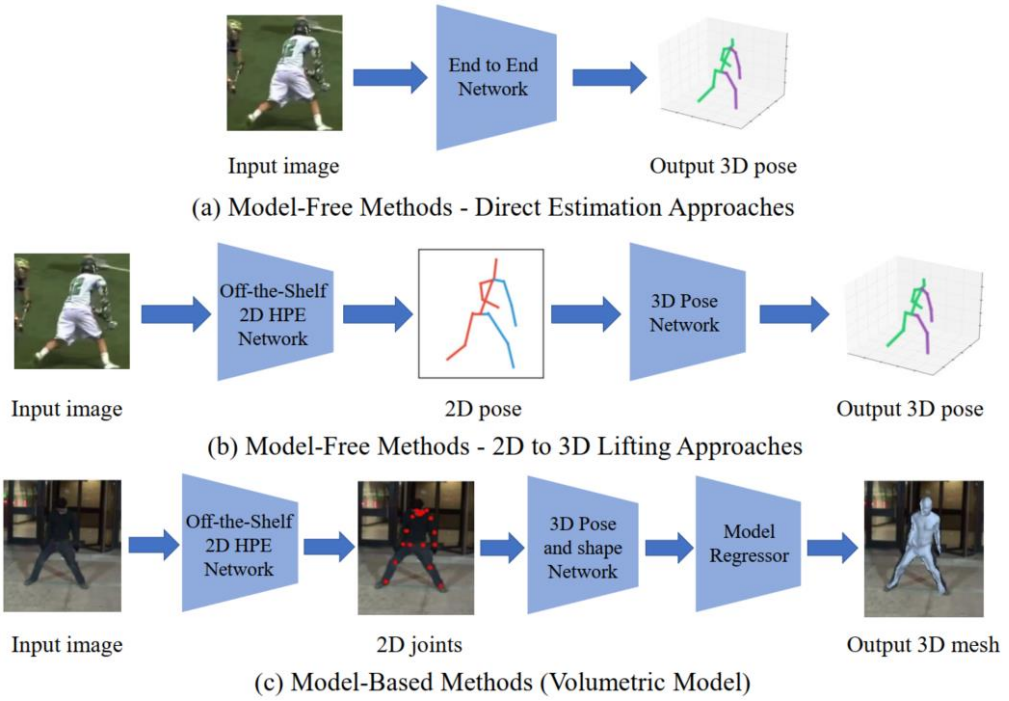

Abbildung 17 | Ablauf der Methoden zur 3D-Posenschätzung einzelner Personen [28]

2.3.2.2 3D - Mehrpersonen - Posenschätzung

Die 3D-Mehrpersonen-Posenschätzung aus Bildern oder Videos, die nur aus einer Perspektive aufgenommen wurden, können wie die 2D-Mehrpersonen-Posenschätzung in Top-Down-Ansätze und Bottom-Up-Ansätze eingeteilt werden. Die Eigenschaften der 2D-Top-Down- und Bottom-Up-Ansätze gelten auch für den 3D-Fall. Abbildung 18 bietet einen Überblick über die beiden Methoden [28].

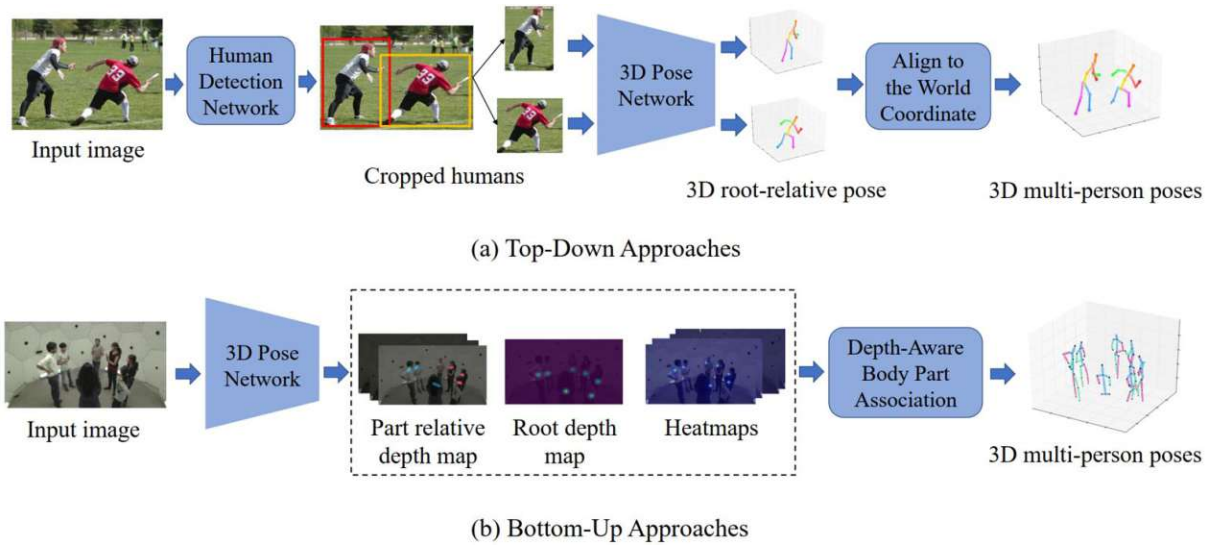

Abbildung 18 | Ablauf der Methoden zur 3D-Posenschätzung mehrerer Personen [28]

Die Top-down-Ansätze der 3D-Posenschätzung für mehrere Personen erkennen zunächst jede einzelne Person. Dann werden für jede erkannte Person eine absolute Wurzelkoordinate (Keypoint im Zentrum des Menschen) und die relative 3D-Pose durch 3D-Posenetzwerke geschätzt. Basierend auf der absoluten Wurzelkoordinate jeder Person und ihrer dazu relativen Pose werden anschließend alle Posen in einem globalen Koordinatensystem ausgerichtet [28].

Im Gegensatz zu den Top-Down-Ansätzen ermitteln Bottom-Up-Ansätze zunächst alle Körpergelenkpositionen inklusive Tiefeninformationen sowie die Wurzelkoordinaten und ordnen dann jeder Person die Körperteile entsprechend der Wurzeltiefe und der relativen Körpergelenktiefe zu. Die wesentliche Herausforderung bei Bottom-Up-Ansätzen besteht darin, die Keypoints zu den jeweiligen Personen zu gruppieren [28].

Top-Down-Ansätze erzielen in der Regel mit den neuesten Methoden zur Erkennung von Personen und zur Schätzung der Pose einzelner Personen vielversprechende Ergebnisse. Der Rechenaufwand und die damit verbundene Analysedauer können jedoch mit zunehmender Anzahl von Menschen übermäßig hoch werden, insbesondere in überfüllten Szenen. Im Gegenteil dazu weisen die Bottom-Up-Ansätze einen linearen Berechnungs- und Zeitaufwand auf. Da die Top-Down-Ansätze zuerst einen Begrenzungsrahmen für jede Person bestimmen und anschließend nur den Bereich innerhalb betrachten, könnten globale Informationen in der Szene vernachlässigt werden. Es besteht die Möglichkeit, dass die geschätzte Tiefe des zugeschnittenen Bereichs nicht mehr mit der tatsächlichen Tiefenordnung übereinstimmt und die geschätzten menschlichen Körper dann in überlappenden Positionen dargestellt werden. Nach dem Erkennen aller einzelnen Personen kann jedoch das menschliche Körpernetz jeder Person leicht wiederhergestellt werden, indem der modellbasierte 3D-Einzelpersonen-HPE-Schätzer integriert wird. Für die Bottom-Up-Ansätze ist es hingegen nicht so einfach, 3D-Körpernetze zu erstellen. Hierfür wird ein zusätzliches Modellregressormodul benötigt, um die menschlichen Körpernetze basierend auf den endgültigen 3D-Posen zu rekonstruieren [28].

3 State-of-the-Art / Literaturanalyse

Dieses Kapitel soll einen Überblick über aktuelle Entwicklungen der menschlichen Posenschätzung und der damit verbundenen Ergonomiebewertung verschaffen. Zuerst werden Methoden zur 2D- und 3D-Analyse vorgestellt. Anschließend folgen aktuelle Anwendungen der Posenschätzung für die Ergonomiebewertung. Zum Abschluss werden zum Vergleich noch andere Methoden vorgestellt, welche die Ergonomie ohne Posenschätzung beurteilen.

3.1 Schätzung der menschlichen Pose

In diesem Kapitel werden aktuelle Algorithmen und Entwicklungen zur 2D- und 3D-Posenschätzung vorgestellt.

3.1.1 2D - Posenschätzung

OpenPose stellte den ersten Ansatz dar, welcher die 2D-Pose von mehreren Personen in einem Bild in Echtzeit erkannt hat. Zuvor war die Posenschätzung mit erheblich längeren Rechenzeiten verbunden. Die Verarbeitung in Echtzeit ist aber eine Schlüsselkomponente, um Maschinen ein Verständnis für Personen in Bildern und Videos zu ermöglichen. Die Methode basiert auf gefalteten neuronalen Netzwerken und verwendet eine nichtparametrische Darstellung, die als Part Affinity Fields bezeichnet wird. Part Affinity Fields sind eine Reihe von 2D-Vektorfeldern, die die Position und Ausrichtung von Gliedmaßen auf dem Bild erkennen. Daraus können die Körperteile zu Personen verknüpft werden. Dieses Bottom-Up-System erzielt unabhängig von der Anzahl der Personen im Bild eine hohe Genauigkeit bei einem Bruchteil der Rechenleistung [12].

Abbildung 19 zeigt die einzelnen Schritte zur Erzeugung der Pose mit OpenPose. Als Eingabe für das gefaltete Netzwerk dient das gesamte Bild (a). Dieses erzeugt zum einen sogenannte Part Confidence Maps (b), welche für jedes Pixel in Form von Heatmaps die Wahrscheinlichkeit angeben, ob sich dort das jeweilige Körpergelenk befindet. Zum anderen werden Part Affinity Fields (c) generiert, die mittels eines 2D-Vektorfelds zwei Gelenke miteinander verbinden. Im darauffolgenden Schritt (d) werden alle möglichen Kombinationen von Gelenkpaaren analysiert und das plausibelste ausgewählt. Diese werden schließlich zu den Posen der Personen auf dem Bild zusammengesetzt (e) [12].

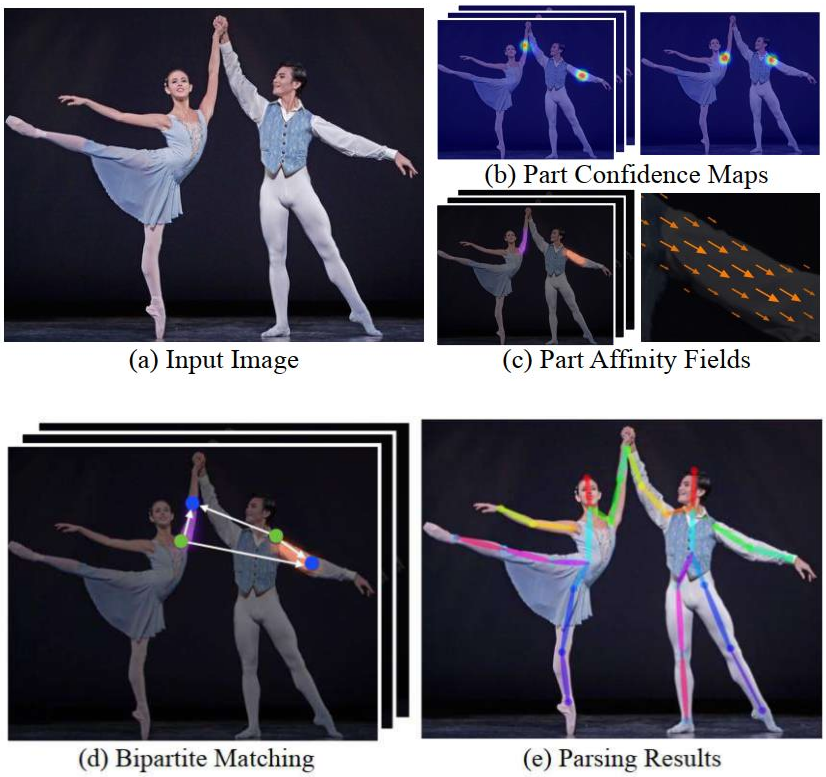

Abbildung 19 | Ablauf von OpenPose [12]

Die Lokalisierung von Keypoints ist selbst für Menschen eine sehr herausfordernde Aufgabe. Die Schwierigkeit resultiert aus der unterschiedlichen Kleidung, der Überschneidung von Gliedmaßen, der Verformung menschlicher Gelenke in verschiedenen Posen und dem komplexen Hintergrund. In jüngster Zeit basieren die meisten modernen Methoden zur Schätzung der menschlichen Pose auf der Heatmap-Regression. Die endgültigen Koordinaten der Keypoints werden durch direktes Decodieren der Heatmaps erhalten. Wang et al. stellte einen Top-Down-Ansatz vor, um genauere Lokalisierungsergebnisse zu erhalten. Dieser realisiert hauptsächlich zwei Verbesserungen: Einerseits wurde zwischen grober und genauer Lokalisierung unterschieden und dafür jeweils verschiedene Merkmale und Methoden angewandt. Andererseits berücksichtigt die Methode auch die Beziehung zwischen den Keypoints [29].

Zunächst findet eine grobe Lokalisierung basierend auf den Kontextinformationen statt. Beispielsweise wird ein Bereich mit Fingern und Armen untersucht, ob sich in der nahegelegenen Umgebung ein Handgelenk-Keypoint befindet. Dieser Schritt kann als Vorschlagsprozess bezeichnet werden. Nach der groben Lokalisierung folgt ein Verfeinerungsprozess, der die Detailstruktur des Handgelenks weiter beobachtet, um die genaue Position des Handgelenk-Keypoints zu bestimmen. Das System verwendet daher zwei verschiedene Subnetze, mit jeweils unterschiedlichen Merkmalskarten für den Vorschlags- bzw. den Verfeinerungsprozess. Diese zweistufige Methode erzielt sowohl hinsichtlich der Effektivität als auch in Bezug auf die Leistung hervorragende Ergebnisse. Abbildung 20 zeigt einen Überblick des Netzwerks [29].

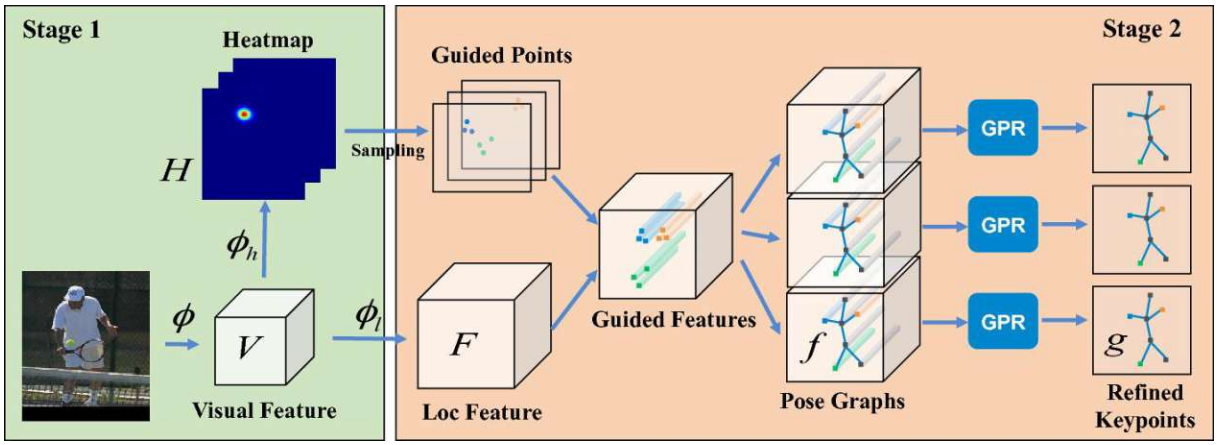

Abbildung 20 | Architektur des zweistufigen Posenschätzungsnetzwerks [29]

In der ersten Stufe wird ein Heatmap-Regressor angewendet, um grobe Lokalisierungs-Heatmaps und daraus wiederum einen Satz von Punkten zu erhalten. In der zweiten Stufe werden aus diesen Punkten mit entsprechenden Lokalisierungsmerkmalen Posendarstellungen erstellt und in ein GPR-Modul (Graph Pose Refinement) eingespeist. Das GPR-Modul führt den Verfeinerungsprozess durch, bei dem die Berücksichtigung der Beziehungen von Keypoints dazu beiträgt, falsche Vorhersagen zu vermeiden und zu korrigieren. Beispielsweise können verdeckte Gelenke aus den Positionen der benachbarten Keypoints mithilfe der Beschränkungen des menschlichen Körpers (Beweglichkeit der Gelenke, Länge der Körperteile) geschätzt werden [29].

Störfaktoren wie Überschneidungen, schnelle Bewegungen und Bewegungsunschärfe erschweren in Videos neben der Schätzung auch das Tracking der Posen. Zusätzlich sind Videodatensätze zum Trainieren der neuronalen Netzwerke weniger vielfältig, wodurch es schwierig ist, robuste Algorithmen zu erzielen. Beim Tracking von Posen besteht das Ziel darin, menschliche Posen in allen Frames eines Videos zu schätzen und sie im Laufe der Zeit den jeweiligen Personen zuzuordnen. Moderne Tracking-Ansätze beruhen auf optischem Fluss oder Personen-Wiedererkennung. Beide Ansätze haben jedoch Nachteile. Wenn eine Person verdeckt ist, schlägt der optische Fluss fehl, weil der Zusammenhang verloren geht. Die Wiedererkennung ermöglicht zwar die erneute Zuordnung von Personen, selbst wenn diese für längere Zeit verschwunden sind, es bleibt jedoch schwierig, teilweise verdeckte Personen mit Top-Down Ansätzen zu erkennen. Rafi et al. stellten daher ein Netzwerk vor, welches Informationen aus dem vorherigen Frame verwendet. Abbildung 21 zeigt die durchgeführten Schritte des Systems. Bei einer gegebenen Folge von Frames bestimmt die Methode eine Reihe von Begrenzungsrahmen für Personen und führt eine Top-Down-Posenschätzung durch. Das Verfahren verwendet zuvor trainierte Zusammenhänge der Keypoints aufeinanderfolgender Frames, um nicht erkannte Posen wiederherzustellen und erkannte bzw. wiederhergestellte Posen einzelnen Personen zuzuordnen. Das gesamte Netzwerk benötigt keine Videodaten für das

Training, da es zum Erlernen der Keypointzusammenhänge mit einzelnen Bildern trainiert wird. Verschwindende Keypoints können auf diese Weise durch Zuschneiden der Bilder simuliert werden. Dadurch ist es robust gegenüber Bewegungsunschärfe und schweren Überschneidungen [30].

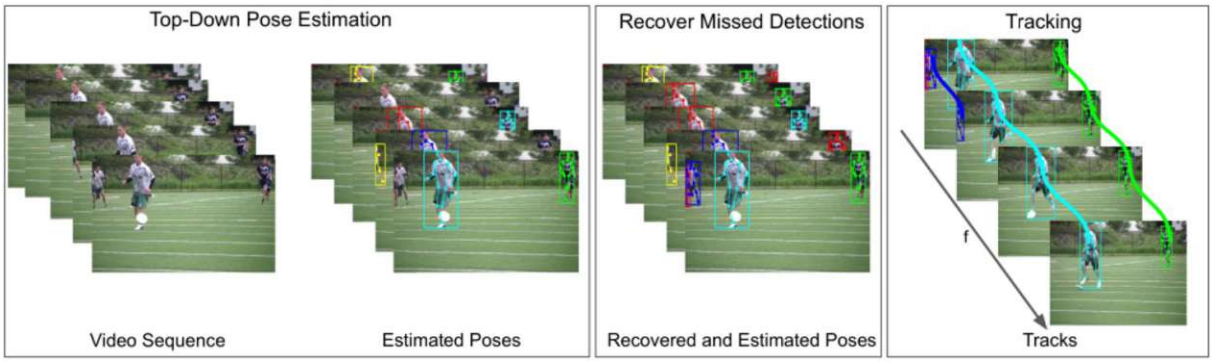

Abbildung 21 | Übersicht der Tracking-Methode [30]

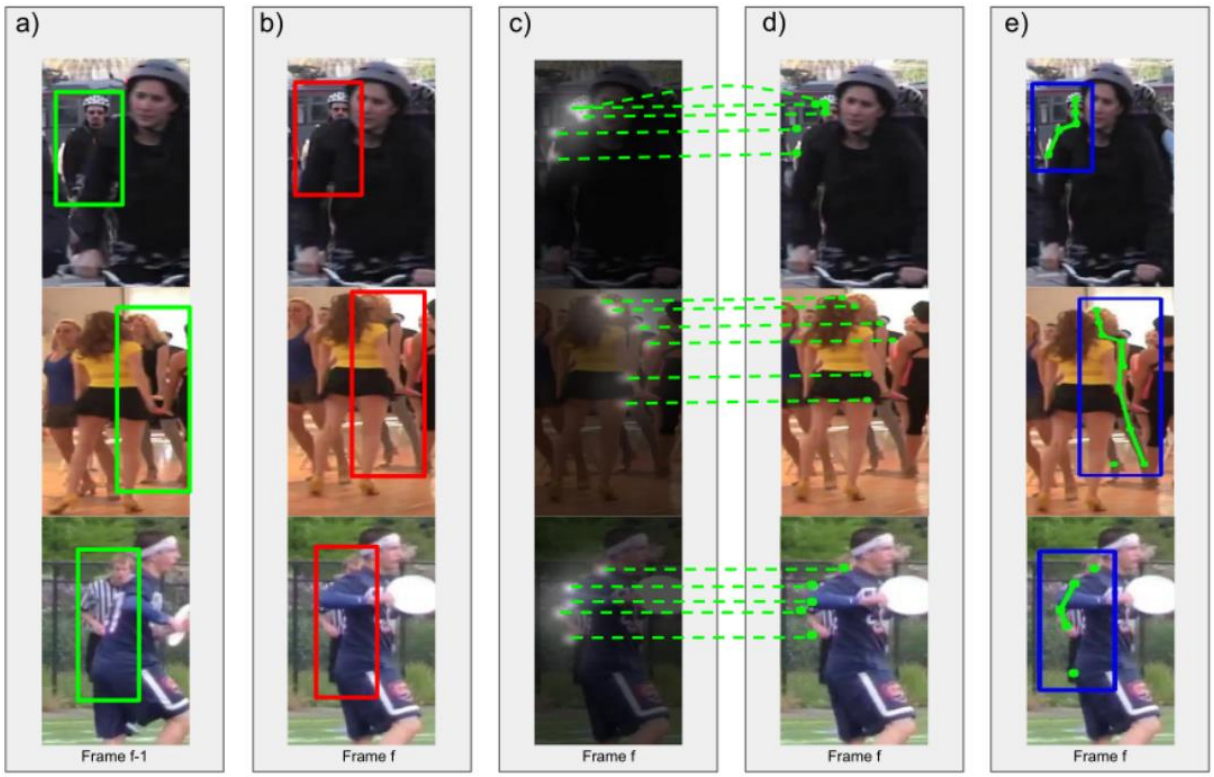

Abbildung 22 | Erkennung von verdeckten Personen [30]

Der Ablauf der Wiederherstellung fehlender Erkennungen ist in Abbildung 22 dargestellt. Punkt (a) auf der linken Seite zeigt eine vom Top-Down-Algorithmus erkannte Person im vorhergehenden Frame f-1. Im darauffolgenden Bild (b) wurde die Person aufgrund einer Überschneidung übersehen. Abschnitt (c) projiziert die Keypoints aus Frame f-1 auf den nächsten Frame f mit der verdeckten Person (d). Daraus können der Begrenzungsrahmen und die entsprechenden Keypoints geschätzt werden (e) [30].

In vielen Anwendungsfällen werden nur die Hauptgelenke des Körpers betrachtet. Jin et al. beschäftigen sich mit der Ganzkörper-Posenschätzung, also der gleichzeitigen Lokalisierung der Keypoints von Körper, Gesicht, Händen und Füßen. Allerdings enthalten vorhandene Datensätze für das Training der neuralen Netzwerke keine Werte des gesamten menschlichen Körpers. Frühere Arbeiten haben daher ihre Modelle separat mit verschiedenen Datensätzen von Gesicht, Hand und menschlichem Körper trainiert. OpenPose kombiniert beispielsweise mehrere tiefe neurale Netzwerke, die unabhängig voneinander mit verschiedenen Datensätzen trainiert wurden, um die Körper-, Gesichts- und Handhaltung zu erkennen. Diese Verfahren können Nachteile haben. Z.B. ist die Größe der aktuellen Datensätze von 2D-Hand-Keypoints begrenzt. Die meisten Ansätze zur Schätzung der Handhaltung verwenden im Labor aufgezeichnete oder synthetische Datensätze, welche die Leistung der vorhandenen Methoden in realen Szenarien beeinträchtigen. Daher wurde ein großer Datensatz für die Ganzkörper-Posenschätzung namens COCO-WholeBody entwickelt, der zusätzlich die Begrenzungsrahmen von Gesicht und Hand sowie die Keypoints von Gesicht, Hand und Fuß enthält. Mit diesem können die hierarchische Struktur des menschlichen Körpers und die Korrelationen zwischen verschiedenen Körperteilen berücksichtigt werden, um die gesamte Körperhaltung abzuschätzen. Dies ermöglichte die Entwicklung eines zuverlässigeren Top-Down-Posenschätzers für den menschlichen Körper inklusive Gesichts-/Handerkennung, Gesichtsausrichtung und 2D-Handposenschätzung [31].

Die gleichzeitige Vorhersage aller Keypoints der Ganzkörperhaltung führt jedoch zu einer schlechteren Leistung, da die Größenordnungen von menschlichem Körper, Gesicht und Hand unterschiedlich sind. Beispielsweise erfordert die Posenschätzung des Körpers einen großen Faltungskern, um Überschneidungen und komplexe Posen handhaben zu können. Die Schätzung der Keypoints von Gesicht und Hand erfordern für eine genaue Lokalisierung jedoch eine höhere Bildauflösung. Wenn alle Keypoints gleich behandelt und direkt vorhergesagt werden, ist die Leistung nicht optimal. Um die Unterschiede bei den Größenordnungen der Ganzkörper-Posenschätzung auszugleichen, entwickelten Jin et al. das Top-Down-Netzwerk ZoomNet. Für jede Person lokalisiert ZoomNet zunächst die leicht zu erkennenden Körper-Keypoints und schätzt die grobe Position von Händen und Gesicht. Anschließend zoomt der Algorithmus in das Bild, um sich auf die Hand- und Gesichtsbereiche zu konzentrieren. Durch die höhere Auflösung ist eine genauere Lokalisierung dieser Keypoints möglich. Abbildung 23 zeigt das Ergebnis dieser Methode. Im Gegensatz zu früheren Ansätzen, bei denen normalerweise mehrere Netzwerke zusammengesetzt werden, verfügt ZoomNet über ein einziges Netzwerk, das durchgängig trainiert werden kann. Umfangreiche Experimente zeigten, dass ZoomNet mit dem Datensatz COCO-WholeBody den Stand der Technik deutlich übertrifft [31].

Abbildung 23 | Ergebnis der Ganzkörper-Posenschätzung mit ZoomNet und COCO-WholeBody [31]

3.1.2 3D – Posenschätzung

Eine Vielzahl von 3D-Posenschätzungsmethoden verfolgt den 2D-to-3D-Lifting-Ansatz, welcher aus der 2D-Pose die 3D-Pose ermittelt. Daher resultieren die Fortschritte bei der 2D-Posenschätzung auch in eine erhebliche Verbesserung der 3D-Posenschätzung [28].

Tome et al. entwickelten die erste Methode, die die 2D-Keypoints nicht als gegeben betrachtet, sondern die 2D- und die 3D-Posenschätzung gemeinsam nutzt, um beide Aufgaben zu verbessern. Die 3D-Informationen sind dabei in einer zusätzlichen Schicht im gefalteten neuronalen Netzwerk eingebettet. Diese erzeugt aus den 2D-Daten die 3D-Pose und stellt sicher, dass die 3D-Pose auch physikalisch plausibel ist. Der Vorteil der Einbeziehung der 3D-Pose ist, dass auch die Lage der 2D-Keypoints verbessert wird, weil sie die anatomischen Einschränkungen des menschlichen Modells erfüllen muss. Auf diese Weise profitieren beide Teile voneinander. Ein weiterer Vorteil dieses Ansatzes besteht darin, dass die 2D- und 3D-Trainingsdatenquellen eigenständig sein können. Durch diese Entkopplung von 2D- und 3D-Trainingsdaten, besteht die Möglichkeit, die Trainingssätze völlig unabhängig voneinander zu erweitern [32].

Abbildung 24 zeigt die mehrstufige tiefe Architektur für die 2D- und 3D-Schätzung der menschlichen Pose. Jede Stufe erzeugt als Ausgabe einen Satz von Heatmaps, die darstellen, mit welcher Wahrscheinlichkeit sich ein Keypoint an einer bestimmten Position befindet. Die Heatmaps sowie das auszuwertende Bild dienen als Input für die nächste Stufe. In den einzelnen Stufen ermittelt ein 2D-Posenschätzer vorläufige Heatmaps, aus denen mithilfe eines Modells eine 3D-Pose bestimmt wird. Diese 3D-Pose wird anschließend wieder in die Ebene projiziert und mit der zuvor berechneten 2D-Pose fusioniert. Dadurch erhöht sich mit jeder Stufe schrittweise die Genauigkeit der Keypoints [32].

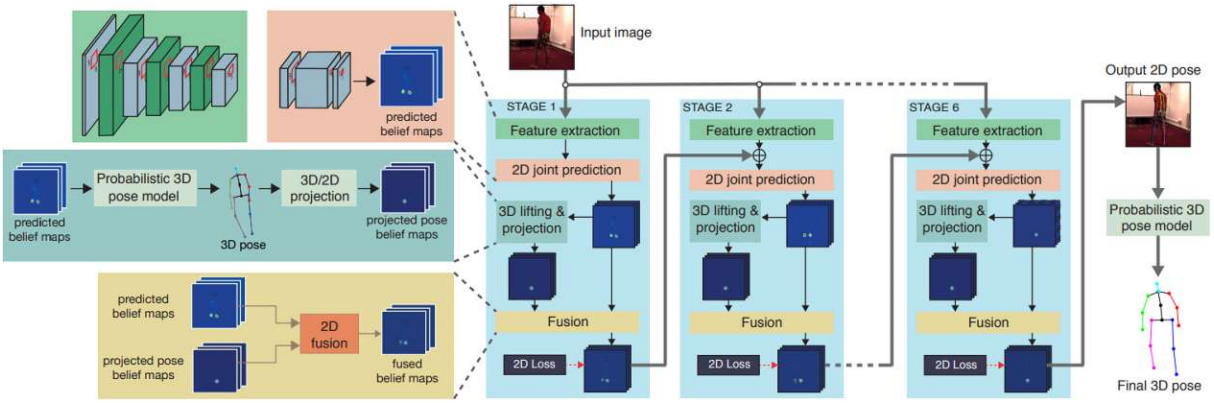

Abbildung 24 | Mehrstufiger tiefer Algorithmus zur gleichzeitigen 2D- und 3D-Posenschätzung [32]

Auch bei der Schätzung der menschlichen Pose in 3D sind Überschneidungen das Hauptproblem. Um mit den Überschneidungen in Videos besser umgehen zu können, entwickelten Cheng et al. ein tiefes Netzwerk, das aus drei Teilen besteht. Das erste Netzwerk gibt die geschätzten 2D-Positionen der Keypoints für eine Person in Form von Heatmaps Frame für Frame aus. Mithilfe eines optischen Flusses der Bildsequenz wird beurteilt, ob die vorhergesagten Keypoints verdeckt sind oder nicht. Andere Methoden verwenden für die anschließende Posenschätzung oft alle Keypoints, obwohl einige davon aufgrund von Überschneidungen sehr ungenau sind. Im Gegensatz dazu werden hier verdeckte Keypoints herausgefiltert und die möglicherweise unvollständigen 2D-Keypoints an den zweiten und dritten Teil weitergegeben. Die beiden nächsten Schritte ermitteln dann die 2D- und die 3D-Pose, wobei auch die Daten aus der Vergangenheit berücksichtigt werden, um kontinuierliche Bewegungen sicherzustellen. Durch die Verwendung von unvollständigen - anstelle von vollständigen, aber falschen - 2D-Keypoints sind die Netzwerke von den fehleranfälligen Schätzungen verdeckter Keypoints weniger betroffen. Für das Training des gesamten Netzwerkes wurde ein Datensatz mit Überschneidungen erstellt, der durch Projektion eines Zylindermodells in verschiedenen Posen aus verschiedenen Betrachtungswinkeln auf die 2D-Ebene realisiert wurde. Abbildung 25 zeigt die unterschiedlichen Ergebnisse der 2D- und 3D-Posen ohne Herausfiltern der verdeckten Keypoints (oben) und mit Herausfiltern der verdeckten Keypoints (unten) [33].

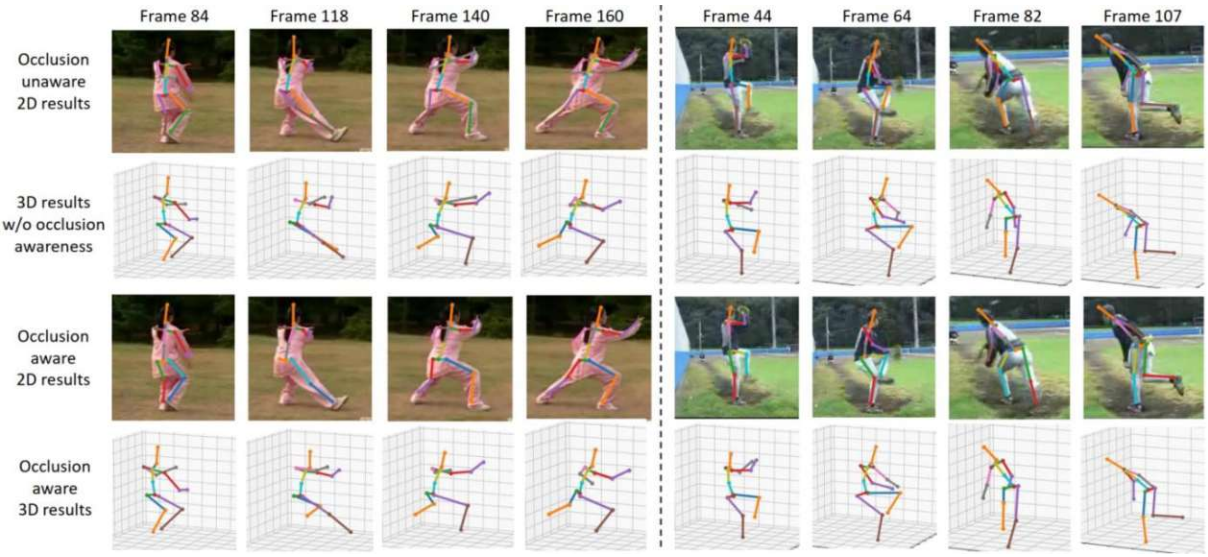

Abbildung 25 | Vergleich der Ergebnisse ohne und mit Herausfiltern verdeckter Keypoints [33]

Die Methoden für die 3D-Posenschätzung einzelner Personen lassen sich nicht einfach für mehrere Personen verallgemeinern, da sie hohe Anforderungen an den Speicher und die Leistung stellen. Dieser Nachteil begrenzt auch die Auflösung der Heatmaps, wodurch Quantisierungsfehler entstehen. Fabbri et al. entwickelten eine einfache und effektive Methode namens LoCO (Learning on Compressed Output), welche hochauflösende volumetrische Heatmaps auf eine kompakte und besser handhabbare Darstellung abbildet. Dies spart Speicher- und Rechenaufwand, während der größte Teil des informativen Inhalts erhalten bleibt. Diese komprimierte Darstellung der Daten ermöglicht die 3D-Posenschätzung für mehrere Personen mit einem Bottom-Up-Ansatz. Das Kernstück der Methode ist der sogenannte Volumetric Heatmap Autoencoder, ein gefaltetes Netzwerk, welches die Aufgabe hat, Heatmaps zu einer dichten Zwischendarstellung zu komprimieren. Ein zweites Modell, der Code Predictor, nutzt diese Daten dann, um die kodierten 3D-Positionen der Keypoints aus einem Bild zu schätzen. Anschließend werden die Daten dekomprimiert, um die ursprüngliche Darstellung wiederherzustellen. Auch in 100 Meter breiten Umgebungen mit mehr als 50 Personen konnten damit vielversprechende Ergebnisse erreicht werden. Die Laufzeit der Methode ist dabei unabhängig von der Anzahl der Personen im Bild. Darüber hinaus ist die Methode durch das hochauflösende Ergebnis auch in der Lage, eine genaue 3D-Pose für eine einzelne Person zu schätzen. Abbildung 26 zeigt beispielhafte Ergebnisse der 3D-Posenschätzung mit LoCO [34].

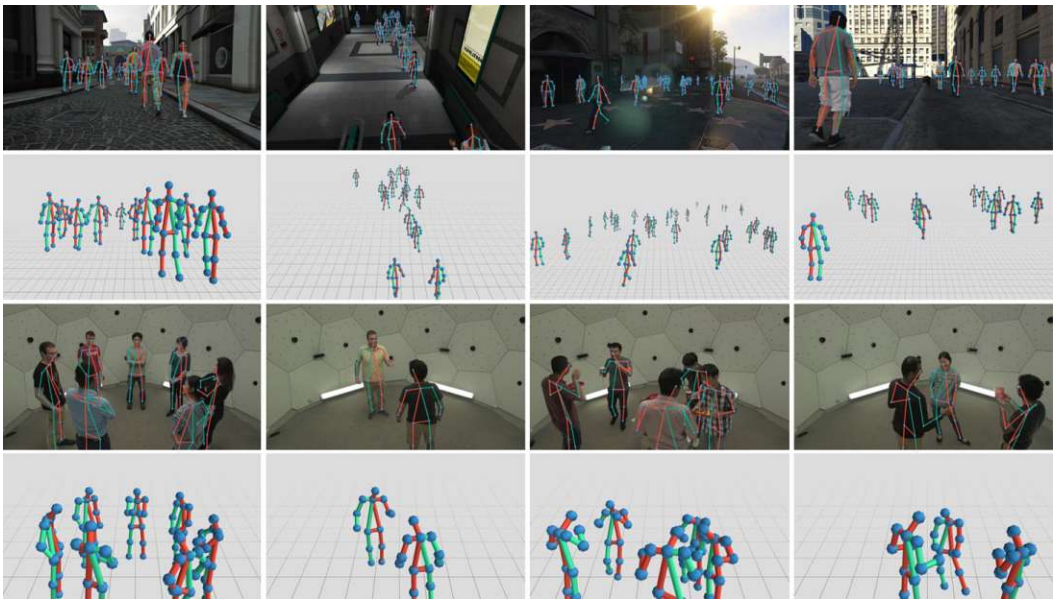

Abbildung 26 | Ergebnisse der 3D-Posenschätzung von mehreren Personen mit LoCO [34]

Benzine et al. entwickelten ein Top-Down-Netzwerk namens PandaNet (Pose estimationAtioN and Dectection Anchor-based Network), welches die Begrenzungsrahmen und die 2D- sowie 3D-Posen mehrerer Personen schätzt. Im Gegensatz zu anderen Top-Down-Mehrpersonen-Ansätzen hängt die Rechengeschwindigkeit nicht von der Anzahl der Personen ab. Somit können Bilder mit einer großen Anzahl an Personen verarbeitet werden. Zusätzlich ist die Methode unempfindlich gegenüber Größenschwankungen und starken Überschneidungen der Personen. Um dies zu erreichen, verwendet das Modell sogenannte Anker. Bei den Ankern handelt es sich um eine Reihe von diskreten vordefinierten Feldern. Das Netzwerk sagt dann voraus, ob ein Anker ein bestimmtes Körperteil enthält oder nicht. Wird das Körperteil erkannt, speichert der Anker seine vollständige 3D-Pose. Diese ankerbasierte Formulierung ermöglicht Ausgaben mit einer geringeren Auflösung als die Heatmap-Formulierung, da ein einzelnes Ausgabepixel ausreicht, um die gesamte Pose einer Person zu speichern. Diese Eigenschaft ist wichtig, um Personen, die sich sehr weit im Hintergrund befinden, effizient zu verarbeiten. Beinhaltet ein Anker Teile von mehreren Personen, wird er vom Algorithmus ausgeschlossen. Abbildung 27 zeigt diesen Vorgang. Links ist das Eingabebild dargestellt. Zuerst wird ein Gitter von Ankern berechnet. Ein Anker ist im zweiten Bild beispielhaft gelb dargestellt. Anschließend erfolgt die Zuweisung der Anker zu den verschiedenen Personen (rot und blau). Die mehrdeutigen Anker mit überlappenden Personen werden dann herausgefiltert [35].

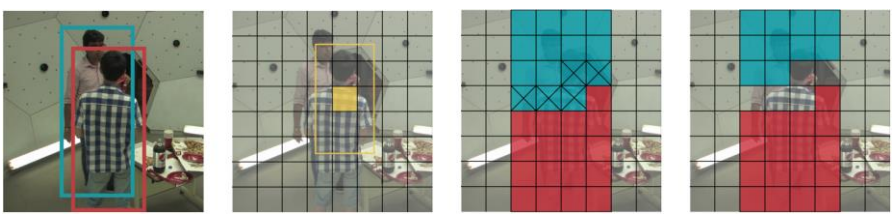

Abbildung 27 | Auswahl der Anker für die Personenerkennung [35]

3.2 Bewertung der Ergonomie mittels Posenschätzung

Paudel und Choi führten ein automatisches Posenschätzungssystem ein, das den Körperwinkel einer Arbeitskraft berechnet und angibt, ob sich die Gelenkwinkel für die Ausführung von Aufgaben in einem sicheren Bereich befinden. Es kombiniert das System zur 2D-Mehrpersonenposenschätzung OpenPose mit Beobachtungsmethoden zur Ergonomiebewertung wie dem Rapid Entire Body Assessment (REBA) und dem Rapid Upper Limb Assessment (RULA). Die REBA-Methode bewertet dabei das Risiko des gesamten Körpers mit einem Score von 1 (minimales Risiko) bis 15 (maximales Risiko). Basierend auf den, mit OpenPose ermittelten, 2D-Positionsdaten der Gelenke, werden die Winkel zwischen den Körperteilen berechnet. Die Methode ermittelt anschließend aus diesen Winkeln den jeweiligen REBA- bzw. RULA-Gesamtscore und somit das Risiko der Arbeitskraft. Bei risikobehafteten Posen ab einem REBA-Score von 6 erzeugt das System automatisch ein Warnsignal. Dadurch können ergonomische von nicht ergonomischen Posen unterschieden werden. Der Ablauf dieser Methode ist in dem Blockdiagramm in Abbildung 28 dargestellt. Das System ist derzeit allerdings nur in der Lage, einzelne Bilder auszuwerten und die Posen werden oft nicht richtig erkannt, wenn einzelne Keypoints für die Kamera nicht ersichtlich sind. Ein großer Nachteil dieser Methode ist zusätzlich, dass sie für die Beurteilung der Körperhaltung nur die 2D-Pose verwendet. Da sich die zweidimensionalen Winkel aber teilweise stark von den realen dreidimensionalen Winkel unterscheiden, lassen sich damit nur bedingt Rückschlüsse auf die Ergonomie ziehen [36].

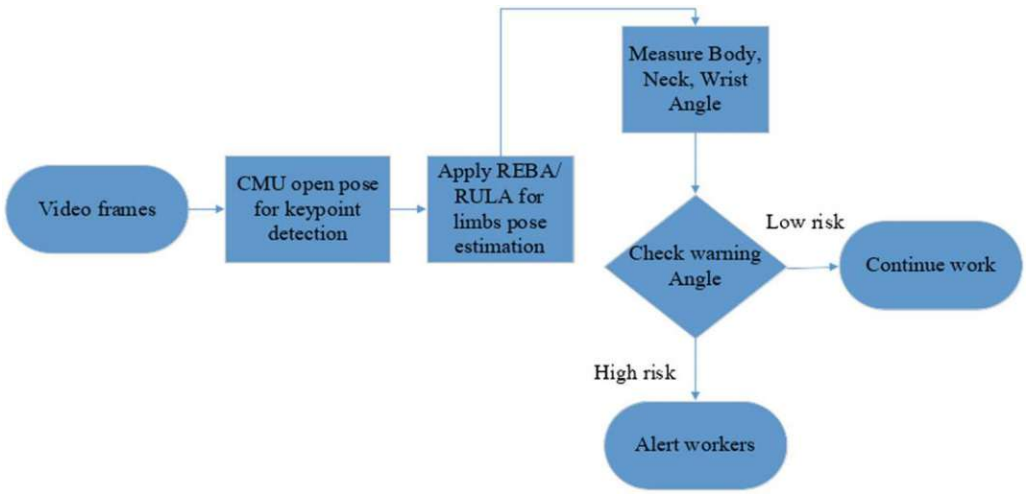

Abbildung 28 | Blockdiagramm des Systems zur Ergonomiebewertung mit Posenschätzung [36]

Altieri et al. lösten dieses Problem durch ein Bewegungsanalysesystem, das auf einem Netzwerk von drei Kameras basiert und somit die Messung verschiedener Winkel für die Beurteilung der Körperhaltung ermöglicht. Die Kameras sind dabei wie in

Abbildung 29 platziert, sodass sie die Arbeitskraft von oben, von vorne und von der Seite erfassen. Die obere Kamera wird hauptsächlich dazu verwendet, die Winkel der Handgelenke und Handbewegungen aufzuzeichnen. Die frontale Ansicht dient der Erfassung des seitlichen Abspreizens und der Drehung der Arme. Um die Körperhaltungen in Bezug auf die Beugung und Streckung der Arme zu bestimmen, ist die seitliche Ansicht nötig. Eine Synchronisation erfolgt unter Verwendung einer Glühbirne, welche fixiert und gleichzeitig jeder Kamera gezeigt wird [37].

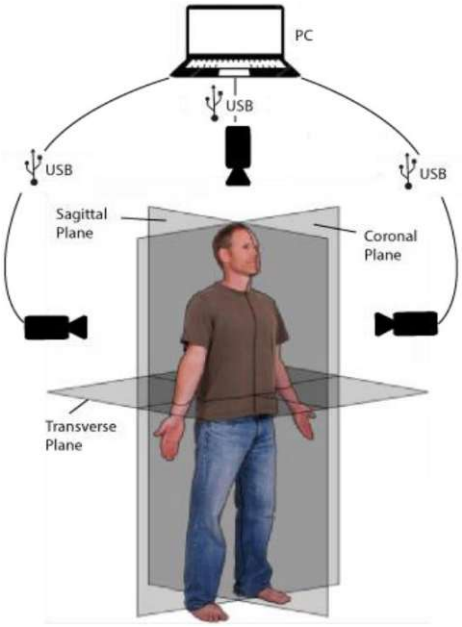

Abbildung 29 | Anordnung der Kameras zur Erfassung des menschlichen Körpers [37]

Um die Körperhaltung der Arbeitskraft während der ausführenden Aufgaben zu erkennen, nutzt das System die zweidimensionale Mehrpersonen-Posenschätzungsmethode OpenPose. Zusätzlich wurde ein Algorithmus erstellt, der die Ausrichtung der wichtigsten Punkte in Bezug auf Schultern und Becken berücksichtigt, um festzustellen, ob das Motiv vor, seitlich oder unter der Kamera positioniert ist. Basierend auf der Orientierung der Person werden die, von OpenPose ausgegebenen, Keypoint-Koordinaten dazu verwendet, die, in der jeweiligen Ansicht sichtbaren, Winkel der Gelenke zu berechnen. Abbildung 30 zeigt ein Beispiel für die Berechnung der Winkel der Armbeugung aus der seitlichen Kameraansicht (links) und des Abspreizens aus der frontalen Ansicht (rechts) [37].

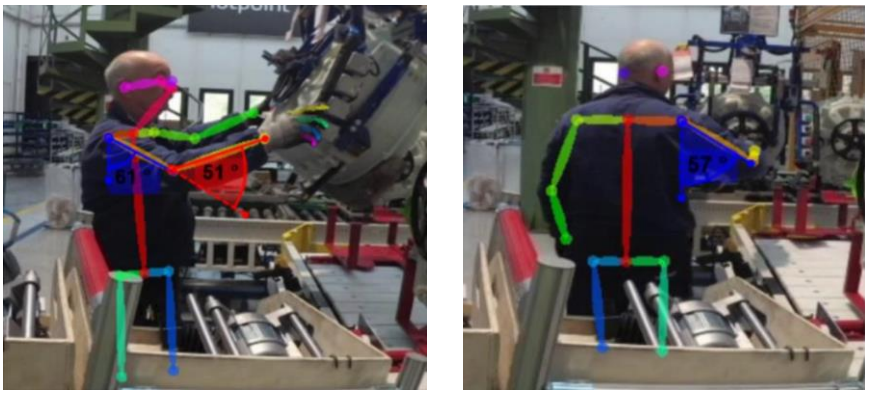

Abbildung 30 | Beispiele für die Ermittlung der Gelenkwinkel aus verschiedenen Ansichten [37]

Um die Körperhaltung zu beurteilen, wurde in dieser Arbeit die OCRA-Methode (Occupational Risk Assessment) ausgewählt. Die OCRA-Checkliste stellt ein Instrument zur Risikobewertung bei sich wiederholenden Aufgaben dar, die durch schnelle und sich wiederholende Bewegungen mit vernachlässigbarer Last gekennzeichnet sind. Dafür wurden für jedes Körperteil die Sekunden ermittelt, in denen es einen nicht ergonomischen Winkel einnimmt. Aus diesen Informationen lässt sich der OCRA-Index berechnen. Zur Validierung wurde das System in einer realen industriellen Fallstudie mit der Bewertung durch eine Ergonomie-Fachkraft verglichen. Die berechneten Werte lagen meist unter denen, die von der Fachkraft ermittelt wurden. Es konnten jedoch keine signifikanten Unterschiede in Bezug auf die Risikoklassenschätzung zwischen den ergonomischen Bewertungen festgestellt werden [37].

Chu et al. entwickelten ein System, welches mit Aufnahmen einer einzelnen Kamera arbeitet. Diese Methode nutzt für die Analyse eine 3D-Posenschätzung. Dabei dienen auch Videos von ganzen Arbeitsvorgängen als Eingangsdaten. Das Resultat ist eine ergonomische Analyse der einzelnen Körperteile bzw. auch eine Bewertung der gesamten Körperhaltung. Das System wurde mit der Programmiersprache Python implementiert. Eine Übersicht ist in Abbildung 31 dargestellt [38].

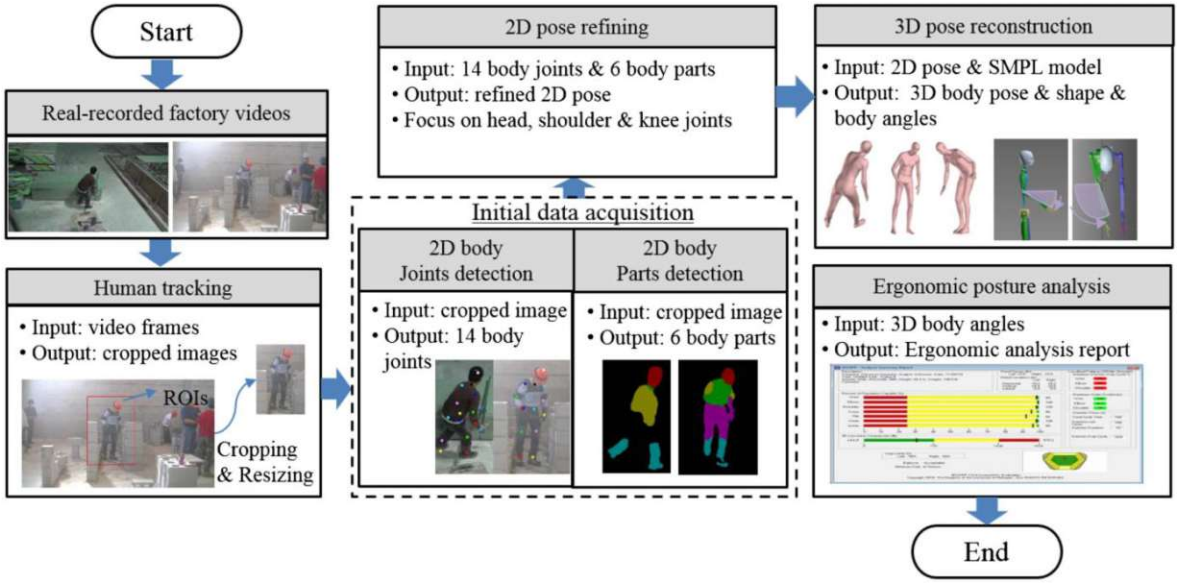

Abbildung 31 | Ablauf der Ergonomieanalysemethode mit indirekter 3D-Posenschätzung [38]

Zunächst wird die Arbeitskraft mit einer Kamera visuell getrackt, um den, für die Analyse relevanten, Bereich abzugrenzen. In der Regel werden mehrere Personen gleichzeitig erfasst. Die Beurteilung der ergonomischen Haltung richtet sich jedoch speziell an eine Arbeitskraft. Das visuelle Tracking basiert auch auf einem gefalteten neuronalen Netz und kann dabei helfen, die jeweilige Person zu verfolgen, zu lokalisieren und von anderen zu unterscheiden. Des Weiteren kann das visuelle Tracking dazu beitragen, irrelevante und ablenkende Hintergrundinformationen, wie z.B. Schreibtische, Fenster, Industriematerialien und andere Menschen, zu entfernen. Auf diese Weise müssen die verbleibenden Arbeitsschritte nur noch auf diesem beschränkten Bereich angewandt werden. Anschließend schätzen ein Gelenk- und ein Körperteilerkennungsalgorithmus die Positionen von 14 Gelenken und sechs Körperteilen der Arbeitskraft. Durch die Unterscheidung zwischen Gelenken und Körperteilen können die 2D-Posen im nächsten Schritt genauer bestimmt werden. Die Idee dabei ist, dass sich die Gelenke in ihren entsprechenden Körperteilen befinden sollen. Somit beschränkt sich die Suche nach den Gelenken eher auf die Bereiche der entsprechenden Körperteile als auf den gesamten menschlichen Körper. Dies führt dazu, dass die Gelenkpositionen genauer erkannt und Verwechslungen (z.B. zwischen linker und rechter Schulter) vermieden werden. Die genaueren 2D-Gelenke werden dann verwendet, um den menschlichen 3D-Körper zu rekonstruieren und die Gelenkwinkel für die Beurteilung der ergonomischen Haltung zu berechnen. Dabei wird das volumetrische Modell auf die Kameraansicht projiziert und seine Parameter angepasst, um den Fehler zwischen dem projizierten Modell und der 2D-Pose immer weiter zu verringern. Am Ende erhält man den 3D-Körper (Gelenke und Form), der optimal auf die Gelenke abgestimmt ist. Daraufhin folgt die Berechnung der Winkel von sieben Gelenken. Jedes Gelenk ist durch zwei Winkel in horizontaler und vertikaler Richtung beschrieben. Die Unterscheidung zwischen horizontaler und vertikaler

Darstellung ermöglicht es, verschiedene Ergonomiebewertungsinstrumente zu verwenden, obwohl sie typischerweise unterschiedliche Anforderungen an die Eingabe der Gelenkwinkel stellen. Eine Übersicht über die horizontalen und vertikalen Winkel ist in Abbildung 32 dargestellt [38].

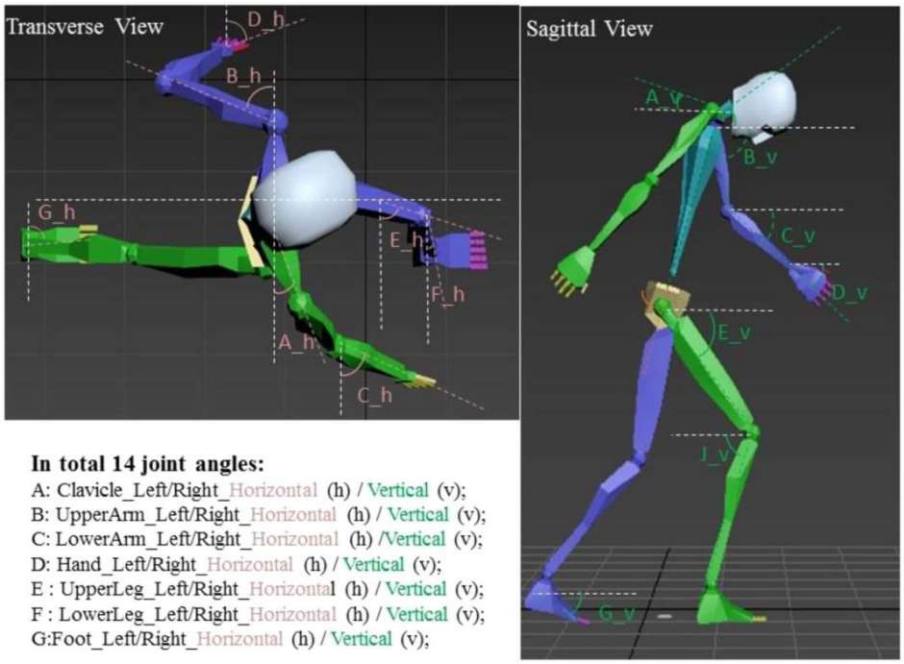

Abbildung 32 | Definition der horizontalen und vertikalen Winkel [38]

Um das System zu testen, wurden die Videos von zwei realen Szenarien verwendet. Der rekonstruierte 3D-Körper stimmte im Allgemeinen mit dem Körper der Arbeitskraft sehr gut überein. Darüber hinaus konnte die Rekonstruktion auch dann durchgeführt werden, wenn die Person teilweise verdeckt war. In einem Szenario trug die Arbeitskraft zusätzlich einen IMU-basierten Motion-Capture-Anzug, um die Genauigkeit des Verfahrens zu bestimmen. Der Vergleich brachte einen durchschnittlichen Fehler von 11° hervor. Abschließend wurde die Ergonomie mit den beiden Methoden 3DSSPP (3D Static Strength Prediction Program) und REBA (Rapid Entire Body Assessment) analysiert. Diese lieferten als Ergebnisse einen Score für jedes Körperteil, der in Diagrammen über die Zeit dargestellt wurde [38].

Die Haltung bestimmt nicht allein das Risiko für körperliche Schäden. Auch die Art der Aufgabe und das Objekt, das an der Aktivität beteiligt ist, haben großen Einfluss darauf. Die Wiederholung bestimmter Bewegungen kann ebenso zu zusätzlichem Druck auf bestimmte Körperteile führen. Hinzu kommt, dass verschiedene Personen eine Aufgabe nicht unbedingt auf die gleiche Weise ausführen. Deshalb haben Parsa und Banerjee einen Ansatz entwickelt, der zusätzlich zur Analyse der Körperhaltung auch die durchgeführte Aktivität erkennt und in Klassen (z.B. Gehen, Heben, Aufnehmen, Platzieren,…) einteilt. Dieser Ansatz vereint beide Prozesse mit einem sogenannten Multi-Task Framework. Dies ist ein einzelnes Netzwerk zur Lösung mehrerer verwandter Aufgaben. In diesen Netzwerken werden die Daten in mehreren

Zweigen verarbeitet, die jeweils für eine bestimmte Aufgabe verantwortlich sind. Die Zweige bestehen dann wiederum aus neuronalen Netzwerken. Normalerweise gibt es eine Hauptaufgabe sowie mehrere Hilfsaufgaben, die die Kernaufgabe ergänzen. Hier besteht die Hauptaufgabe darin, die REBA-Ergebnisse vorherzusagen. Die Hilfsaufgabe führt die Klassifizierung der Tätigkeiten durch. Abbildung 33 zeigt das grundsätzliche Schema dieser Methode [39].

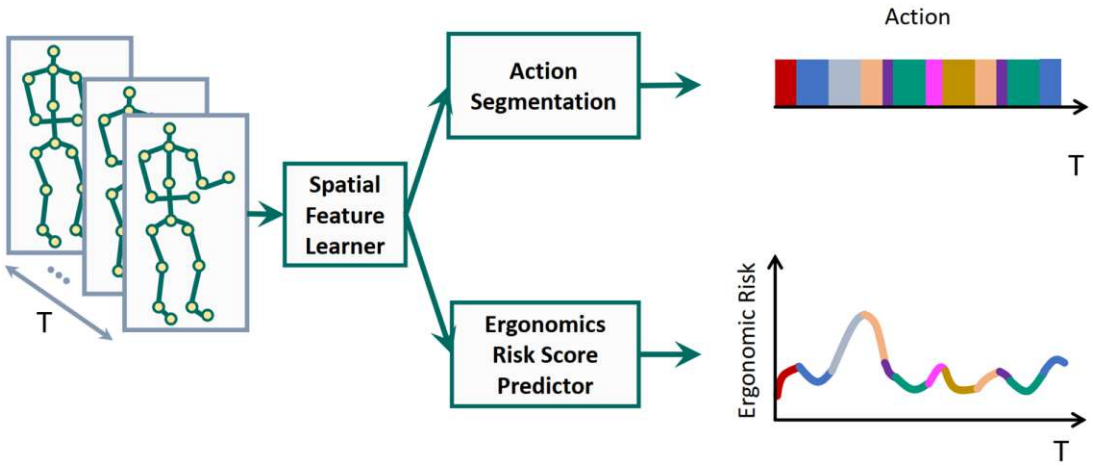

Abbildung 33 | Multi-Task-Aktivitätsklassifizierung und Ergonomie-Risikobewertung [39]

Als Eingabeinformationen dienen bei diesem Multi-Task-Modell die 3D-Gelenkpositionen einer längeren Videosequenz. Eine auf gefalteten Netzwerken basierende Struktur extrahiert aus diesen Daten räumliche Merkmale, welche anschließend für die Aufgabenklassifizierung und die Risikobewertung verwendet werden. Die Aufgabenklassifizierung hat die Funktion, die Aktivitäten zu identifizieren und die entsprechenden Anfangs- und Endbilder zu bestimmen. Die Ergebnisse werden in verschiedenfarbigen Streifen dargestellt, wobei jeder Farbe eine Aktivitätsklasse zugeordnet ist. Die Risikobewertung der Ergonomie ordnet den räumlich-zeitlichen Merkmalen einen Score nach der REBA-Methode zu. Dieser wird in einem Diagramm über die Zeit abgebildet. Die Kurve ist dabei in den Farben der zugehörigen Aktivitätsklassen eingefärbt [39].

3.3 Bewertung der Ergonomie ohne Posenschätzung

Caputo et al. entwickelten eine IMU-basierte Methode, welche die Ergonomie an einem industriellen Arbeitsplatz beurteilt. Die Inertial Measurement Units bestehen hierbei aus einem dreiachsigen Beschleunigungsmesser, einem dreiachsigen Gyroskop und einem dreiachsigen Magnetometer. Diese Sensoren stellen den minimalen Satz von Instrumenten dar, um die Lage eines Starrkörpersystems abzuschätzen. Zusammen mit einem Einplatinencomputer (Raspberry Pi) zur Datenverarbeitung und einem Akku für die Stromversorgung sind die IMUs die Hauptkomponenten der Hardware. Um die Bewegungen der Arbeitskraft aufzuzeichnen werden die IMUs - wie in Abbildung 3 - an Becken, Rumpf, Oberarm,

Unterarm, Oberschenkel und Unterschenkel angebracht. Der Einplatinencomputer berechnet dann aus den aufgezeichneten Signalen die Winkel zwischen den erfassten Körperteilen. Zur Visualisierung der Ergebnisse dienen mehrere Diagramme, in denen die Winkel von verschiedenen Körpergelenken über der Zeit aufgetragen sind. Weiters ermittelt ein Algorithmus, welcher mit der Programmiersprache MATLAB entwickelt wurde, aus den berechneten Gelenkwinkel den Score der Ergonomieanalysemethode European Assembly Work-Sheet (EAWS). Die EAWS-Methode ist wie die RULA-Methode ein Beobachtungsverfahren, bei dem durch das Auswerten von Tabellen ein Score gebildet wird [19].

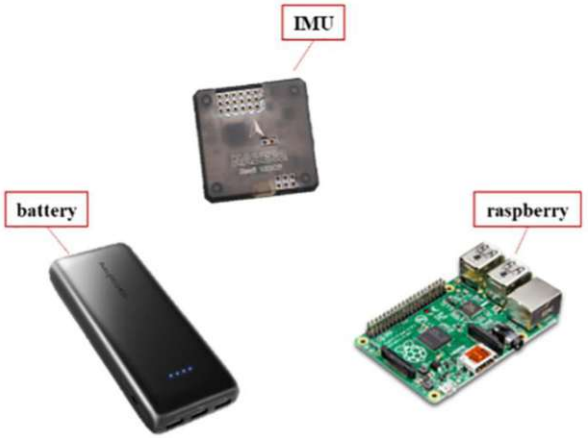

Abbildung 34 | Hardwarekomponenten für die IMU-basierte Ergonomieanalyse [19]

Neben der Ergonomie ist auch die Produktivität an einem Montagearbeitsplatz ein wichtiger Faktor. Um die Produktionsprozesse zu überwachen, zu analysieren und sukzessive zu optimieren, kann die Virtualisierung manueller Vorgänge eine hervorragende Möglichkeit darstellen. Daher entwickelten Ferrari et al. eine indirekte Messmethode namens Motion Analysis System, um manuelle Herstellungs- und Montageprozesse mithilfe von Motion-Capture-Technologien zu überwachen und zu bewerten. Dabei wird ein Netzwerk von Tiefenkameras verwendet, welches die Bewegungen und Körperhaltungen der Arbeitskraft während der Produktionsaktivitäten ohne die Verwendung von Markern erfasst. Eine Softwarearchitektur nutzt diese Daten, um die zeitlichen und räumlichen Aspekte der durchgeführten Tätigkeiten der Arbeitskraft zu analysieren. Dabei werden die Gehwege, Handbewegungen, Kommissionierorte und allgemein der genutzte Raum der Person innerhalb des Arbeitsplatzes z.B. mittels einer Heatmap visualisiert [11].

Die Hardware besteht aus bis zu vier Tiefenkameras, die jeweils mit einem PC verbunden sind. Die PCs kommunizieren über ein WLAN-Netzwerk, um die Positionen der Körpergelenke im überwachten Bereich eindeutig zu bestimmen. Ein genauer Kalibrierungsprozess ist notwendig, um den Standort jeder Kamera in der 3D-Umgebung zu bestimmen. Des Weiteren müssen die Kameras synchronisiert werden, um gleichzeitig die Bewegungen der Arbeitskraft innerhalb des Arbeitsbereichs zu

erfassen. Die Kamerakalibrierung und -synchronisation ermöglicht eine Genauigkeit bei der Lokalisierung der Körperteile von 5-6 cm [11].

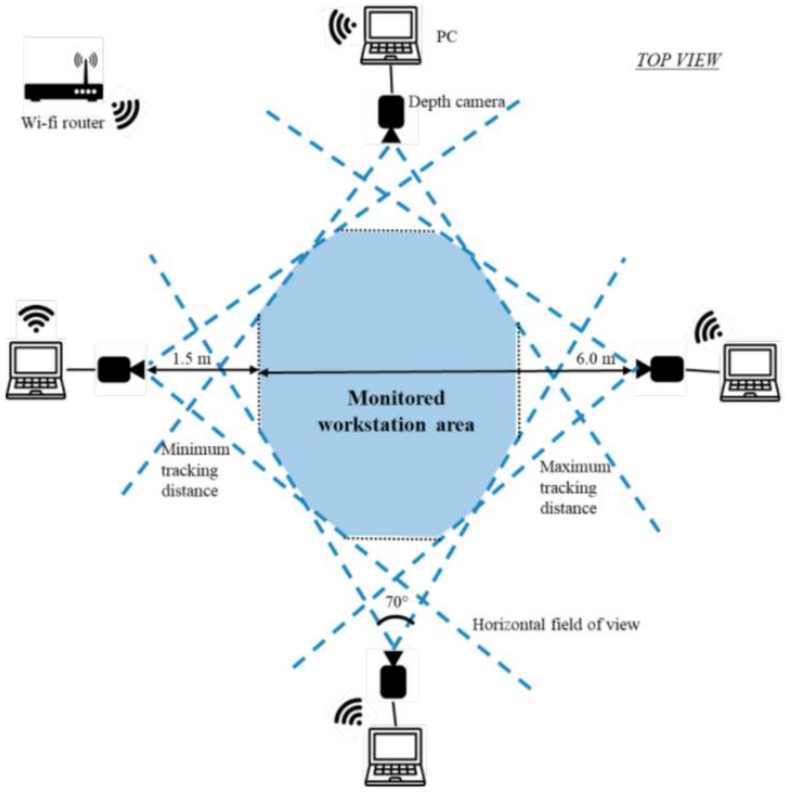

Abbildung 35 | Optimale Konfiguration der Hardwarearchitektur [11]

Die in MATLAB programmierte Softwarearchitektur integriert die Informationen zu den Bewegungen und Körperhaltungen der Arbeitskraft in das digitale Arbeitsplatz-Layout, um automatisch eine Reihe von Kennzahlen zu berechnen. Die Software benötigt dafür Daten wie die physischen Merkmale der Person und die dreidimensionalen Abmessungen und Positionen von Arbeitsplätzen, Werkzeugen und Produkten. Daraus berechnet die Software unter anderem die Pfade und Geschwindigkeiten von verschiedenen Körperteilen und unterscheidet die durchgeführten Aktivitäten zwischen wertschöpfenden (Aufgabenausführung) und nicht wertschöpfenden Tätigkeiten (Gehen, Kommissionieren usw.). Die Ergebnisse können dann in verschiedenen Diagrammen bzw. in Heatmaps wie in Abbildung 36 dargestellt werden [11].

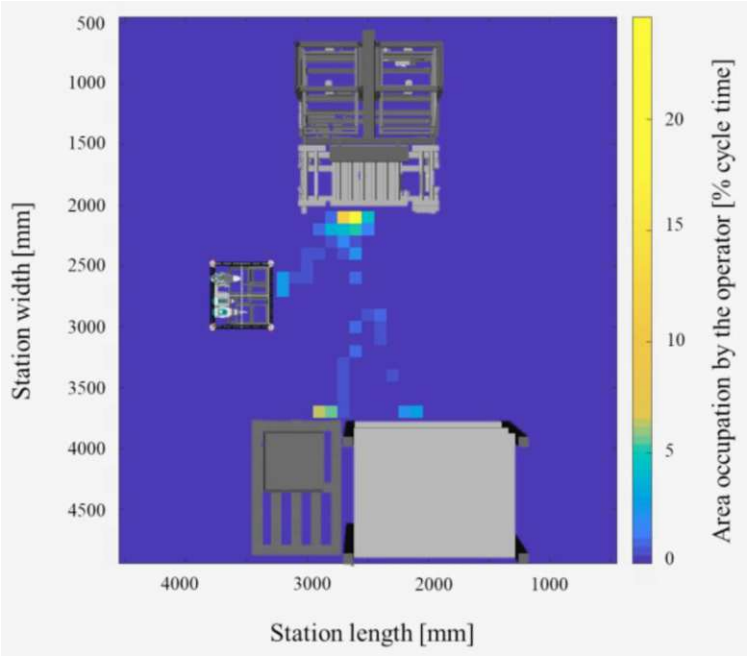

Abbildung 36 | Darstellung der Aufenthalte der Arbeitskraft mittels Heatmap [11]

Kim et al. verglichen Gelenkwinkelmessungen von Marker-, IMU- und Tiefenkamera-basierten Bewegungserfassungssystemen, insbesondere für Fälle mit Überschneidungen von Körperteilen. Für die drei verschiedenen Motion-Capturing-Systeme wurde zum einen ein optisches System von Vicon mit acht Infrarotkameras und 53 reflektierenden Körpermarkern verwendet, als zweites kam das IMU-basierte System Xsens mit 17 Trägheitssensoren zum Einsatz. Als drittes wurde das auf einer Tiefenkamera basierende System Microsoft Kinect auf einem Stativ montiert und in 1,3m Höhe und 2,5m Entfernung positioniert, wobei es genau auf die Vorderseite der Testperson gerichtet wurde. Zur Analyse der drei verschiedenen Methoden dienten, die - in Abbildung 37 dargestellten - sechs statischen Körperhaltungen mit unterschiedlichem Überschneidungsgrad [40].

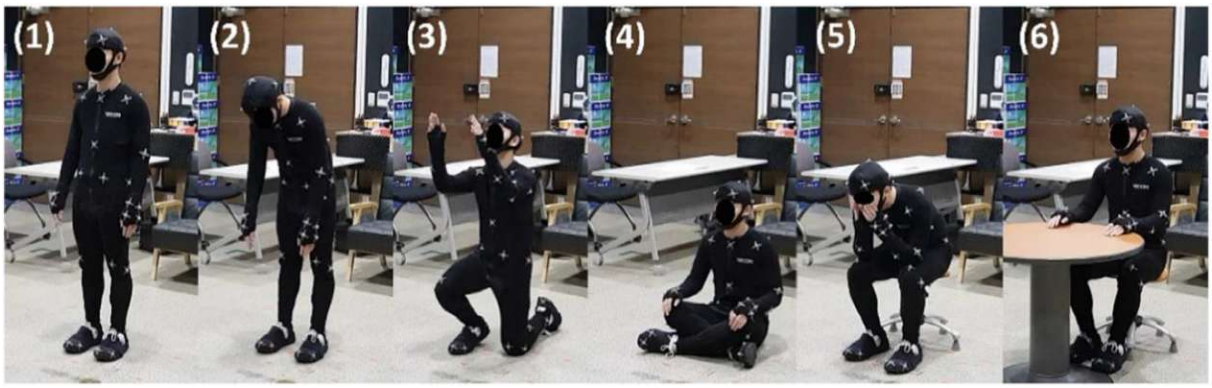

Abbildung 37 | Die getesteten statischen Haltungen: (1) aufrecht stehend, (2) Rumpfbeugung, (3) kniend über dem Kopf arbeiten, (4) mit gekreuzten Beinen auf dem Boden sitzend, (5) Ellbogen auf den Knien und (6) am Schreibtisch [40]

Für RULA- und REBA-Analysen sind 22 Gelenkwinkel erforderlich. Unter diesen wurden 5 Gelenkwinkel (Halsverdrehung sowie Beugung und Verdrehung des linken

bzw. rechten Handgelenks) von der Analyse ausgeschlossen, da Kinect für diese nicht in der Lage war, genaue Messungen durchzuführen. Aus den erhaltenen Positionsdaten der Gelenke wurden die, für die ergonomische Analyse erforderlichen, Winkel berechnet und die mittlere absolute Differenz zwischen den Ergebnissen von Vicon, Xsens und Kinect verglichen. Die Ergebnisse zeigten, dass die Gelenkwinkeldifferenz zwischen den drei Systemen im Allgemeinen gering war. Bei den Fällen mit starken Überschneidungen war das auf der Tiefenkamera basierende System Kinect jedoch weniger stabil als die anderen beiden Bewegungserfassungssysteme [40].

4 Erstellung des Programmcodes

Im Rahmen dieser Diplomarbeit wurde ein Algorithmus entwickelt, der aus einem Video eines Arbeitsvorgangs eine Heatmap erzeugt, welche die Ergonomie an dem Arbeitsplatz abbildet. Abbildung 38 zeigt eine Übersicht über den Ablauf. Ein 3D-Posenschätzer berechnet mittels einer 2D-to-3D-Lifting-Methode die 2D- und 3D-Koordinaten der Arbeitskraft in jedem Frame. Aus diesen Koordinaten werden anschließend Winkel zwischen den einzelnen Körperteilen berechnet und gemäß der RULA-Methode ein Ergonomie-Score gebildet. Dieser Score stellt ein Maß für die Ergonomie der Person im aktuellen Frame dar und reicht von 1 (akzeptable Körperhaltung) bis 7 (gefährliche Körperhaltung). Der Score wird dann mit der aktuellen Position der Arbeitskraft im Bild verknüpft. Dadurch lassen sich Heatmaps erstellen, die die Ergonomie an dem Arbeitsplatz beschreiben. Zusätzlich wurde ein Tracking-Algorithmus implementiert, mithilfe dessen es möglich ist, unbeteiligte Personen herauszufiltern. In diesem Kapitel werden die Anforderungen und der Ablauf des gesamten erstellten Programmcodes näher beschrieben.[1]

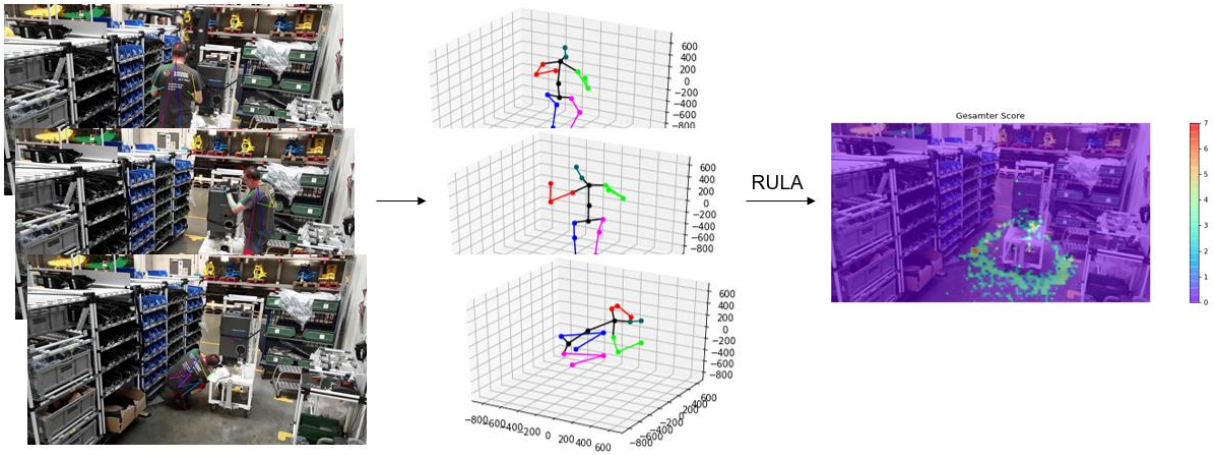

Abbildung 38 | Ablauf der implementierten Methode

4.1 Programmierumgebung

Als Programmierumgebung zur Entwicklung dieses Algorithmus wurde Google Colaboratory (oder kurz Colab) ausgewählt. Google Colab ist eine cloudbasierte virtuelle Umgebung zur Entwicklung und Ausführung von Python-Skripten und ermöglicht es, den Code über eine Browser-basierte Umgebung auszuführen. Die Plattform bietet kostenlosen Zugriff auf Rechenressourcen wie Grafikprozessoren (GPUs) und Tensorprozessoren (TPUs) von Google. Die Nutzung dieser Ressourcen auf den Cloud-Servern von Google verkürzt die Rechenzeit um ein Vielfaches. Somit ist es für EntwicklerInnen möglich, unabhängig von der Hardware des eigenen

[1] Der Code ist auf der GitLab Webseite des Instituts für Managementwissenschaften der TU Wien öffentlich verfügbar.

Computers, rechenintensive Algorithmen wie Deep-Learning-Aufgaben sehr schnell durchzuführen. Google Colab verwendet, wie in Abbildung 39 dargestellt, eine interaktive Programmierumgebung (Jupyter-Notebook), in der gleichzeitig Zellen mit ausführbarem Python-Code und Zellen mit Texten, Überschriften und Bildern erstellt werden können. Die Codezellen sind einzeln ausführbar und geben die Ergebnisse direkt darunter aus. Viele wichtige Python-Bibliotheken wie NumPy für die einfache Handhabung von Vektoren und Matrizen sowie Matplotlib für mathematische Darstellungen sind bereits installiert. Auch im Bereich des maschinellen Lernens verfügt Google Colab über einige Bibliotheken wie z.B. TensorFlow oder PyTorch. Die Dokumente werden in der Cloud Google Drive gespeichert und können somit sehr einfach geteilt und von mehreren Personen bearbeitet werden. Für die Benutzung der Services ist ein Google Account notwendig [41].

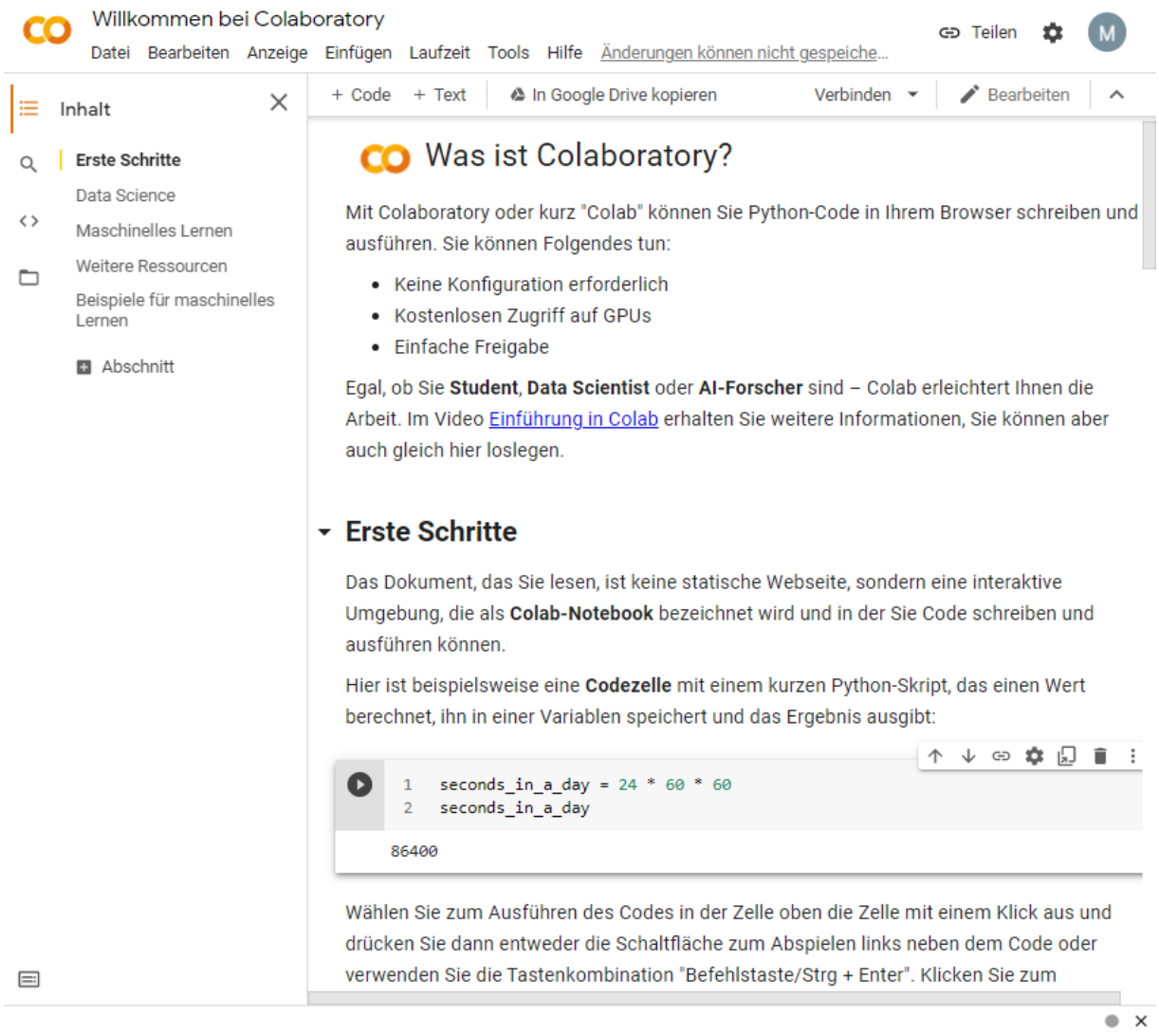

Abbildung 39 | Screenshot aus dem Willkommens-Notebook von Google Colab [41]

Der wichtigste Grund für die Nutzung von Google Colab bei dieser Diplomarbeit war die kostenlose Nutzung von GPUs auf den Cloud-Servern von Google, wodurch bei der Ausführung des Codes erheblich Zeit gespart wurde. Die einfache und

unkomplizierte Nutzung ohne notwendiger Installationen erleichterte den Anfang, da sofort mit dem Programmieren begonnen werden konnte. Da in der Wacker Neuson Linz GmbH, in der die Diplomarbeit geschrieben wurde, jede Installation eine Freigabe erfordert, konnte dadurch zusätzlicher Aufwand eingespart werden. Der Datenaustausch erfolgt über die Cloud-Plattform Google Drive. Das zu analysierende Video muss dorthin hochgeladen werden, um in Google Colab verfügbar zu sein. Die Ergebnisse sind dann wiederum auch in Google Drive verfügbar.

4.2 Posenschätzung

Für die Schätzung der 2D- und der 3D-Pose wurde der, in Kapitel 3.1.2 vorgestellte, Algorithmus von Tome et al. verwendet. Als wichtigstes Kriterium für die Auswahl eines geeigneten Posenschätzers galt die öffentliche Verfügbarkeit (Open Source) und somit die Möglichkeit zum Downloaden des prätrainierten Netzwerks. Der Posenschätzer „Lifting from the Deep“ von Tome et al. erfüllte diese Anforderung. Der Algorithmus benötigt zur Nutzung die Bibliotheken TensorFlow und OpenCV, welche beide in Google Colab bereits vorinstalliert sind [32].

TensorFlow ist eine Open-Source-Bibliothek von Google Brain, der Deep-Learning-Abteilung von Google LLC, welche sich vor allem für das Entwickeln und Trainieren von tiefen neuronalen Netzwerken eignet. Sie ist sehr flexibel einsetzbar und kann beispielsweise für Spracherkennung, Computer Vision und Robotik verwendet werden. Google selbst nutzt TensorFlow z.B. in den Produkten Google Fotos, Google Suche, Google Maps und Google Street View [42] [43].

OpenCV ist eine, von Intel entwickelte, Open-Source-Computer-Vision-Bibliothek zur Extrahierung und Verarbeitung von Daten aus Bildern. Damit ist es möglich, Objekte aufzufinden und zu erkennen, bewegte Objekte zwischen aufeinanderfolgenden Bildern zu tracken und die 2D- oder 3D-Form von Objekten aus einem oder mehreren Bildern zu bestimmen. Die gewonnenen Bilddaten können unter anderem dazu genutzt werden, sie mit einer Kategorie zu verknüpfen. Eine winkende Handbewegung kann so beispielsweise als „Auf Wiedersehen“ interpretiert werden. OpenCV eignet sich sehr gut für Anwendungsfelder wie Gesichts- und Gestenerkennung, Mensch-Computer-Interaktion und mobile Roboter [44].

Das verwendete neuronale Netzwerk zur Posenschätzung von Tome et al. ist bereits trainiert auf der Plattform GitHub verfügbar und damit sofort einsetzbar [45]. Das Netzwerk für die 2D-Posenschätzung wurde mit dem MPI-Datensatz trainiert. Dieser Datensatz umfasst rund 25.000 Bilder von über 40.000 Personen mit den dazugehörigen Körpergelenkpunkten. Die Bilder umfassen mehr als 800 alltägliche menschliche Aktivitäten wie Radfahren oder Tanzen und wurden aus YouTube-Videos extrahiert [46].

Das Netzwerk für die 3D-Posenschätzung wurde mit dem Human3.6M-Datensatz trainiert. Dieser Datensatz besteht aus 3,6 Millionen menschlichen 3D-Posen. Fünf weibliche und sechs männliche Personen führten dafür 15 alltägliche Tätigkeiten wie Gehen, Telefonieren, Essen und Sitzen durch und wurden dabei aus vier verschiedenen Winkeln mit einem Motion-Capture-System aufgenommen [47].

Der Posenschätzer „Lifting from the Deep“ von Tome et al. benötigt als Input ein Bild. Daher wird das zu analysierende Video zuerst mithilfe der Klasse „VideoCapture“ von OpenCV in seine einzelnen Frames zerlegt. Videos haben meist Frameraten von 30 Bildern in der Sekunde. Da sich diese Bilder aber nur geringfügig voneinander unterscheiden, wird hier nur jedes 15. Bild analysiert (entspricht zwei Bildern in der Sekunde). Dadurch kann die Rechenzeit um ein Vielfaches verkürzt werden, ohne wichtige Informationen zu verlieren. Ein Bild wird dem trainierten neuronalen Netzwerk übergeben, welches anschließend sowohl die X- und Y-Koordinaten der erkannten 2D-Posen als auch die X-, Y- und Z-Koordinaten der erkannten 3D-Posen ausgibt. Die 2D-Koordinaten entsprechen dabei der Anzahl der Pixel vom linken bzw. oberen Rand. Bei den 3D-Koordinaten handelt es sich um keine räumlichen Koordinaten, sondern um Koordinaten im Kamerakoordinatensystem (Camera space). Somit können z.B. die Gelenkwinkel gut abgeleitet werden, nicht aber die Körpergröße oder die relative Position zum Tisch. Eine Pose hat die Form einer Liste, bei der jeder Eintrag die X- und Y- bzw. die X-,Y- und Z-Koordinate des jeweiligen Keypoints enthält. Die 2D- und 3D-Posen jedes Frames werden dann wiederum in einer jeweils eigenen Liste gesammelt, um in den nachfolgenden Schritten auf alle Posen des gesamten Videos zurückgreifen zu können. Ein Element in der Liste entspricht also allen erkannten Posen in dem jeweiligen Frame. Werden in einem Bild mehrere Personen erkannt, so beinhaltet der Eintrag in der Liste auch mehrere Posen. Da die Schätzung der Posen mehrere Stunden dauert, werden die Posendaten und die einzelnen Frames zusätzlich auch in Google Drive gespeichert. Dies ermöglicht, spätere Analysen ohne die aufwändige Posenschätzung durchzuführen. Die Daten können dann aus Google Drive wieder importiert werden.

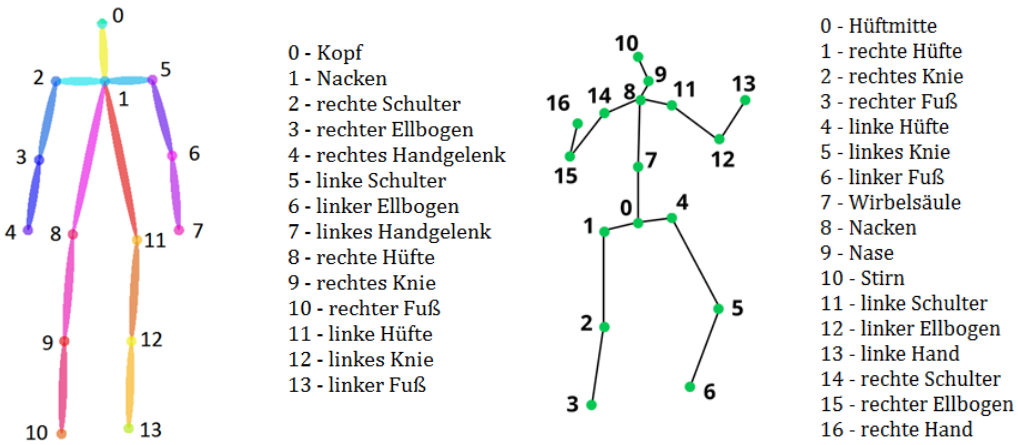

Abbildung 40 | Ausgegebene 2D- und 3D-Koordinaten des Posenschätzers [48] [49]

Eine 2D-Pose stellt sich aus folgenden 14 Keypoints zusammen: Kopf, Nacken, rechte Schulter, rechter Ellbogen, rechtes Handgelenk, linke Schulter, linker Ellbogen, linkes Handgelenk, rechte Hüfte, rechtes Knie, rechter Fuß, linke Hüfte, linkes Knie und linker Fuß. Eine 3D-Pose beinhaltet zusätzlich noch Hüftmitte, Wirbelsäulenmitte und Nase. In Abbildung 40 sind die erhaltenen Posen mit ihren Keypoints dargestellt. Mittels OpenCV ist es möglich, die erhaltenen Koordinaten direkt im Bild als Punkte darzustellen und mit Linien zu verbinden. Die 3D-Posen können in einem dreidimensionalen Diagramm ebenso durch Punkte und Verbindungslinien veranschaulicht werden. Abbildung 41 zeigt das Ergebnis des Posenschätzers anhand eines Beispielbildes.

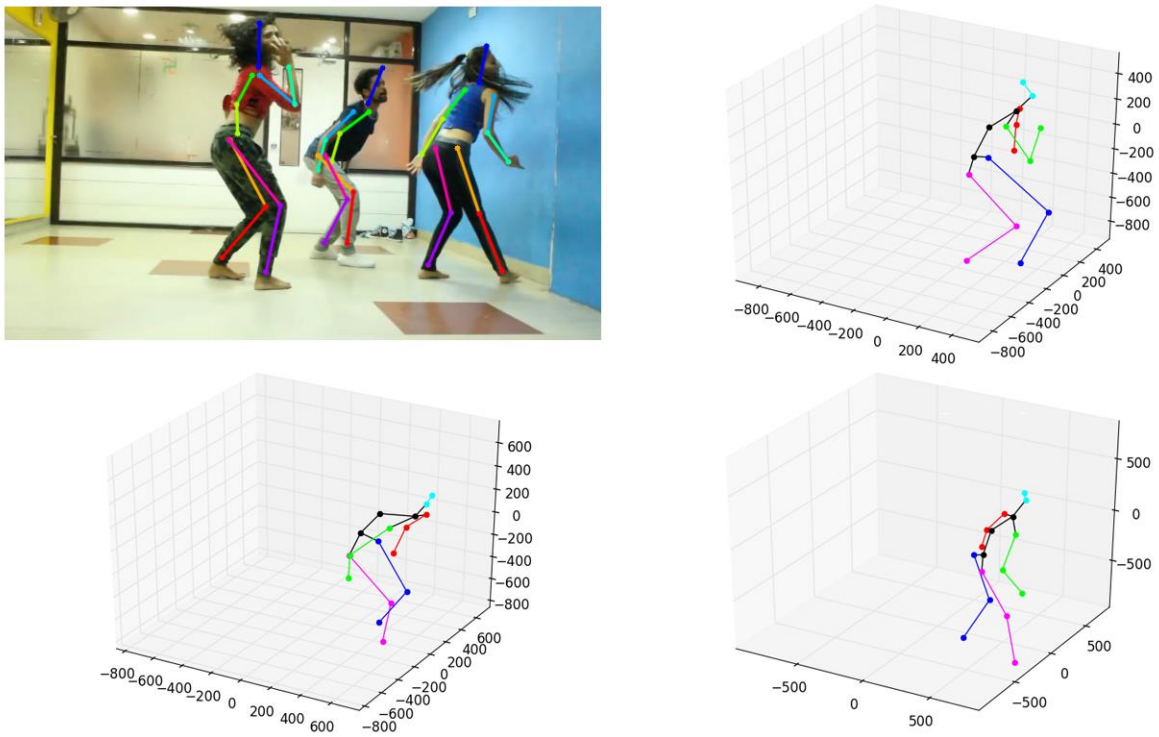

Abbildung 41 | Beispielhafte Ausgabe des Posenschätzers [45]

4.3 Berechnung des Ergonomie-Scores

Die Bewertung der Ergonomie basiert zum Großteil auf der, bereits in Kapitel 2.1.2.2 beschriebenen, RULA-Methode. Das Rapid Upper Limb Assessment ist ein international weit verbreitetes und oft verwendetes Verfahren. Als Ergebnis erhält man einen Score zwischen 1 und 7, welcher sich gut in einer Heatmap darstellen lässt. Für die Beurteilung der einzelnen Körperteile sind meistens bestimmte Bereiche von Winkeln vorgegeben, die sich auch sehr gut für eine automatisierte Berechnung eignen. Aus diesen genannten Gründen wurde die RULA-Methode für die Anwendung in dieser Diplomarbeit ausgewählt.

Um die dafür benötigten Winkel zu berechnen, werden die einzelnen Körperteile als Vektoren zwischen zwei Punkten im Raum dargestellt. Die Punkte folgen aus der Liste der 3D-Posen, welche im vorigen Schritt (Posenschätzung) ermittelt wurde. Durch koordinatenweises Subtrahieren von zwei Punkten lässt sich so ein Vektor aufstellen. Subtrahiert man beispielsweise die X-,Y- und Z-Koordinaten des Handgelenks jeweils mit den X-,Y- und Z-Koordinaten des Ellbogens erhält man einen Vektor vom Ellbogen zum Handgelenk, welcher den Unterarm beschreibt. Auf diese Weise lassen sich alle relevanten Körperteile beschreiben. Der Winkel zwischen zwei Körperteilen entspricht dann dem Winkel, der von zwei Vektoren aufgespannt wird. Dieser berechnet sich, wie in der Formel von Abbildung 42 beschrieben, aus dem Arkuskosinus einer Division mit dem Skalarprodukt der Vektoren als Zähler und der Multiplikation der Beträge der Vektoren als Nenner. Wie in Abbildung 42 dargestellt, existieren zwischen zwei Vektoren immer zwei Winkel. Da sich die Werte der Arkuskosinus-Funktion auf den Bereich zwischen 0° und 180° beschränken, ermittelt die Formel immer den kleineren der beiden Winkel.

$$\theta = \cos^{-1}\left(\frac{\vec{u}\cdot\vec{v}}{|\vec{u}|\cdot|\vec{v}|}\right) = \\ = \cos^{-1}\left(\frac{u_x\cdot v_x + u_y\cdot v_y + u_z\cdot v_z}{\sqrt{u_x^2+u_y^2+u_z^2}\cdot\sqrt{v_x^2+v_y^2+v_z^2}}\right)$$

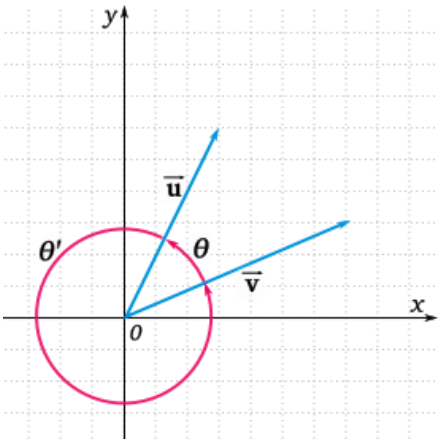

Abbildung 42 | Winkel zwischen zwei Vektoren [50]

Die Berechnung der Winkel erfolgt unter Verwendung der vorher erstellten Liste mit den 3D-Posen aller Frames. Diese Liste wird mittels For-Schleifen durchlaufen. In jedem Schleifendurchgang werden die benötigten Winkel aus den Koordinaten der jeweiligen Pose berechnet und wiederum in Listen gespeichert, um sie später weiterverwenden zu können. Somit existiert für jede Art von Winkel eine Liste, die denselben Winkel in jedem Frame beinhaltet. Das bedeutet, eine Liste enthält alle Ellbogenwinkel des rechten Armes, eine Liste enthält alle Schulterwinkel des linken Armes und so weiter. Die Listen der Winkel werden im nächsten Schritt für die Score-Berechnung benötigt. Um die Scores für die jeweiligen Körperteile zu ermitteln, werden diese Winkellisten wieder mittels For-Schleifen durchlaufen. If-Statements ermöglichen es dann, die Winkel in bestimmte Kategorien einzuteilen und somit zu bewerten. Wie die Winkel und die Posen, werden auch die Scores in Listen gespeichert, um anschließend bei der Heatmap-Erstellung zur Verfügung zu stehen.

4.3.1 Arm-Score

Der Arm-Score beschreibt die Haltung von Oberarm, Unterarm und Handgelenk. Da bei der Posenschätzung das Handgelenk der äußerste Keypoint des Armes ist, kann dafür kein Winkel berechnet werden. Somit ergibt sich der Arm-Score hier nur aus dem Ober- und dem Unterarm-Score. Der Handgelenk-Score wurde in der Berechnung des gesamten Scores auf Eins gesetzt.

Für den Oberarm weist die RULA-Methode, wie in Abbildung 43 dargestellt, jedem Winkelbereich einen Wert von eins bis vier zu. In der Programmierung wurde jeweils mit einem If-Statement überprüft, in welchem dieser Bereiche der aktuelle Winkel liegt. Eine Unterscheidung zwischen der Bewegung nach vorne und der Bewegung nach hinten kann hier nicht realisiert werden, da für beide Fälle derselbe Winkel ausgegeben wird. Dies stellt aber kein Problem dar, da ohnehin beide Varianten einen Wert von zwei ergeben. Für die Berechnung des Oberarm-Winkels wurde als Referenz die senkrechte Z-Achse genommen und nicht der Oberkörper, da sonst im Falle des Bückens ein Winkel entstehen würde - wenn die Arme nach unten hängen. Somit entspricht dies dem Winkel zwischen dem Vektor von der Schulter zum Ellbogen und dem Einheitsvektor in der Z-Richtung (0,0,1). Der sich ergebende Winkel ist dann bei nach oben gestreckten Armen 0°. Um die geforderte Form im RULA-Arbeitsblatt zu erreichen, wurde der berechnete Winkel anschließend von 180° abgezogen.

Für eine Entscheidung, ob die Schultern angehoben sind oder nicht, ist die Genauigkeit der Posenschätzung zu gering. Ebenso kann nicht erkannt werden, ob der Arm unterstützt wird oder ob die Person sich anlehnt. Im Gegensatz dazu kann die Abduktion, also das seitliche Abspreizen des Arms, sehr wohl beurteilt werden. Der Abduktionswinkel beschreibt dabei den Winkel zwischen der Schulterachse und dem Oberarm. Bei hängendem oder nach vorne gestrecktem Arm ist dieser 90°. Als Grenzwert für den Abduktionswinkel wurde 135° gewählt. Bei Überschreiten vergrößert sich der Oberarm-Score um eins. Sollte sich die Haltung der beiden Arme unterscheiden, wird die weitere Berechnung mit dem höheren Score durchgeführt.

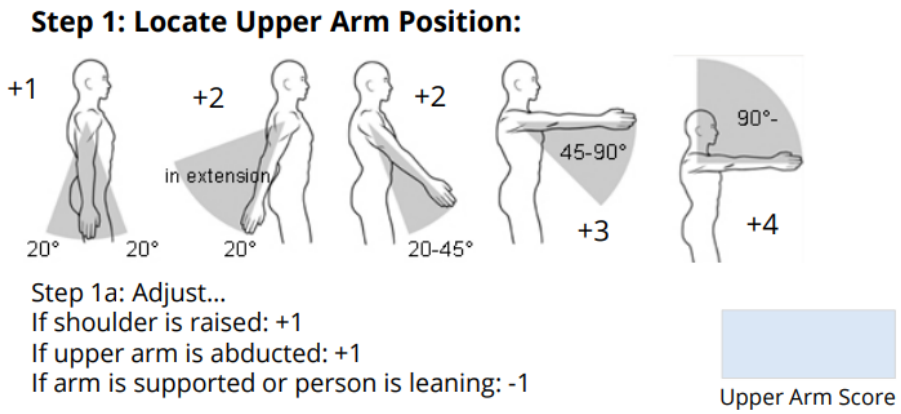

Abbildung 43 | Oberarm-Score [18]

Der Unterarm-Score unterscheidet, wie in Abbildung 44 ersichtlich, nur zwei Abstufungen von Winkeln. Auch hier wird mittels If-Statements überprüft, in welchem Bereich der aktuelle Winkel zwischen Ober- und Unterarm liegt. Diese Definition des Unterarm-Winkels unterscheidet sich gegenüber dem RULA-Arbeitsblatt in der Hinsicht, dass bei ausgestrecktem Arm der Winkel 180° beträgt und nicht 0°. Um mit der RULA-Methode übereinzustimmen, muss der berechnete Winkel wieder von 180° abgezogen werden.

Um zu kontrollieren, ob außermittiges Arbeiten vorliegt, wurden zwei verschiedene Kriterien definiert. Für die Beurteilung der Bewegung nach außen, wird ähnlich zur Abduktion des Oberarms der Winkel zwischen der Schulterachse und dem Unterarm berechnet. Als Grenze wurden hier 100° definiert, um die Ungenauigkeit der Posenschätzung auszugleichen. Überschreitet also der Winkel zwischen Schulterachse und Unterarm einen Wert von 100°, erhöht sich der Unterarm-Score um eins. Die Bewegung über der Körpermitte kann mit diesem Winkel nicht beschrieben werden, da ein Abwinkeln des Ellbogens bei Abduktion des Oberarmes auch einen Winkel kleiner als 90° ergeben würde. Daher wurden jeweils die Abstände von einem Handgelenk zu beiden Schultern berechnet und verglichen. Der Abstand ergibt sich dabei als Betrag des Vektors von dem Handgelenk zur Schulter. Ist zum Beispiel die Distanz vom rechten Handgelenk zur linken Schulter kleiner als zur rechten Schulter, befindet sich die Hand über der Körpermitte. In diesem Fall erhöht sich auch der Unterarm-Score. Wie beim Oberarm, wird auch der Arm mit dem höheren Score für die weitere Berechnung verwendet.

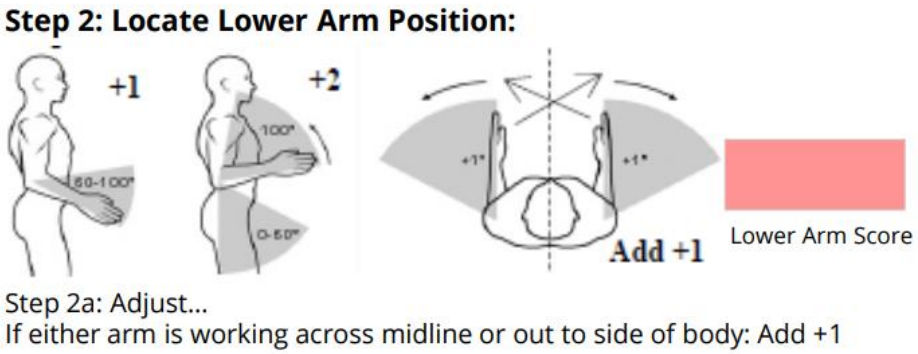

Abbildung 44 | Unterarm-Score [18]

Da sowohl der Handgelenk-Score als auch die Handgelenk-Drehung nicht beurteilt werden können, lässt sich der Arm-Score nur aus Ober- und Unterarm berechnen. Unter der Annahme, dass das Handgelenk keine außerordentlich schädlichen Haltungen einnimmt, ergibt sich das Ergebnis für den Arm-Score aus der ersten Spalte der Tabelle A aus dem RULA-Arbeitsblatt (hier Tabelle 2). Dafür wurde die Spalte in den Programmcode übertragen. Mit den vorher berechneten Werten lässt sich dann der richtige Index ermitteln, um den Arm-Score herauszufinden.

Scores

Table A		Wrist Score							
		1		2		3		4	
Upper Arm	Lower Arm	Wrist Twist		Wrist Twist		Wrist Twist		Wrist Twist	
		1	2	1	2	1	2	1	2
1	1	1	2	2	2	2	3	3	3
	2	2	2	2	2	3	3	3	3
	3	2	3	3	3	3	3	4	4
2	1	2	3	3	3	3	4	4	4
	2	3	3	3	3	3	4	4	4
	3	3	4	4	4	4	4	5	5
3	1	3	3	4	4	4	4	5	5
	2	3	4	4	4	4	4	5	5
	3	4	4	4	4	4	5	5	5
4	1	4	4	4	4	4	5	5	5
	2	4	4	4	4	4	5	5	5
	3	4	4	4	5	5	5	6	6
5	1	5	5	5	5	5	6	6	7
	2	5	6	6	6	6	7	7	7
	3	6	6	6	7	7	7	7	8
6	1	7	7	7	7	7	8	8	9
	2	8	8	8	8	8	9	9	9
	3	9	9	9	9	9	9	9	9

Tabelle 2 | Tabelle A im RULA-Arbeitsblatt zur Ermittlung des Arm-Scores [18]

4.3.2 Nacken-Rumpf-Bein-Score

Der Nacken-Rumpf-Bein-Score beschreibt, wie der Name schon sagt, die Haltung von Nacken, Rumpf und Beinen. Wie beim Arm-Score werden zunächst die einzelnen Werte für Nacken, Rumpf und Beine berechnet. Anschließend wird über eine Tabelle ein gesamter Score ermittelt.

Wie in Abbildung 45 ersichtlich, unterscheidet der Nacken-Score drei verschiedene Neigungen nach vorne und die Neigung nach hinten (Überstrecken genannt). Für das Neigen des Kopfes nach vorne wurde der Winkel zwischen der oberen Wirbelsäule und dem Kopf berechnet. Dies entspricht dem Winkel zwischen dem Vektor vom Nacken- zum Wirbelsäulenmitte-Keypoint und dem Vektor vom Nacken- zum Kopf-Keypoint. Mittels If-Statements wird anschließend überprüft, in welchem Bereich der berechnete Winkel liegt und der jeweilige Score vergeben. Da dieser Winkel bei geradem Kopf genau 180° beträgt und somit bei jeder Neigung des Kopfes - egal in welche Richtung - kleiner wird, benötigt man für die Beurteilung des Überstreckens des Kopfes einen weiteren Winkel. Dafür wurde der Winkel zwischen oberer Wirbelsäule und dem Vektor vom Nacken- zum Nase-Keypoint berechnet. Als Grenzwert wurden hier 140° definiert. Bei Überschreiten dieser 140° liegt Überstrecken des Kopfes vor und der Nacken-Score wird auf den Wert 4 gesetzt.

Zur Überprüfung, ob eine Drehung des Kopfes vorliegt, wurde der Winkel zwischen der Schulterachse und der Nacken-Nase-Linie verwendet. Dieser beträgt bei geradem Kopf 90°. Die Drehung des Kopfes verkleinert bzw. vergrößert den Winkel. Als

Grenzwerte wurden 75° bzw. 105° definiert. Ist der Winkel außerhalb dieses Bereichs, wird der Kopf also um mehr als 15° gedreht, erhöht sich der Nacken-Score um eins. Die seitliche Neigung des Kopfes wird mit dem Winkel zwischen Schulterachse und Nacken-Kopf-Linie beurteilt. Dieser beträgt, so wie der Drehwinkel, bei geradem Kopf 90° und verkleinert bzw. vergrößert sich bei der seitlichen Neigung. Für das If-Statement wurden die gleichen Grenzwerte wie beim Drehwinkel ausgewählt. Ist der Neigungswinkel also kleiner als 75° oder größer als 105°, wird der Nacken-Score um einen Punkt erhöht.

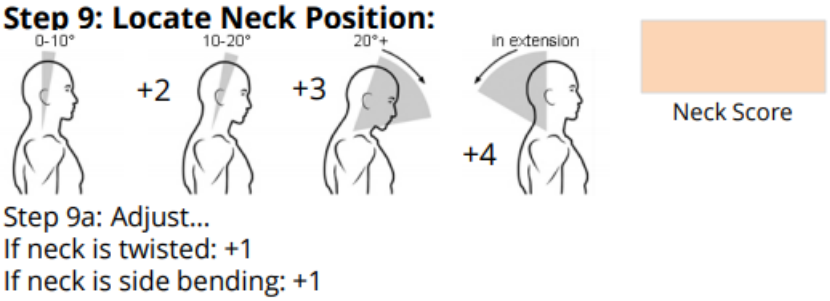

Abbildung 45 | Nacken-Score [18]

Beim Rumpf-Score (dargestellt in Abbildung 46) existieren zur Bewertung der Neigung des Oberkörpers vier verschiedene Abstufungen. Dafür wurde der Winkel zwischen der oberen Wirbelsäule und der senkrechten Z-Achse definiert. Der Winkel zwischen oberer Wirbelsäule und Oberschenkel eignet sich nicht zur Bewertung der Oberkörperneigung, da zum Beispiel im Falle des Sitzens auch ein Winkel von 90° entstehen würde.

Für die Beurteilung der Rumpfdrehung kommt der Winkel zwischen Hüft- und Schulterachse zum Einsatz. Dieser wird aber nur in der XY-Ebene, also in der Ansicht von oben, ausgewertet, da sonst eine Drehung nicht von einer seitlichen Neigung des Oberkörpers unterschieden werden könnte. Im dreidimensionalen Raum würden beide Bewegungen einen Winkel zwischen Hüft- und Schulterachse hervorrufen. Für die zweidimensionale Berechnung eignet sich die Formel in Abbildung 42 ebenso. Die Vektoren werden dann nur mit ihren X- und Y-Koordinaten aufgeschrieben und der Z-Anteil ignoriert. Als Grenzwert für die Rumpfdrehung wurden 15° definiert. Bei Überschreiten des Grenzwerts erhöht sich der Rumpf-Score um eins. Für die Bewertung der seitlichen Neigung des Oberkörpers wurde der Winkel zwischen Hüftachse und oberer Wirbelsäule definiert. Bei geradem Oberkörper beträgt dieser Winkel 90°. Als Grenzwerte wurden wiederum 75° bzw. 105° gewählt. Befindet sich der Neigungswinkel des Oberkörpers außerhalb dieses Bereichs, erhöht sich der Rumpf-Score ebenfalls um einen Punkt.

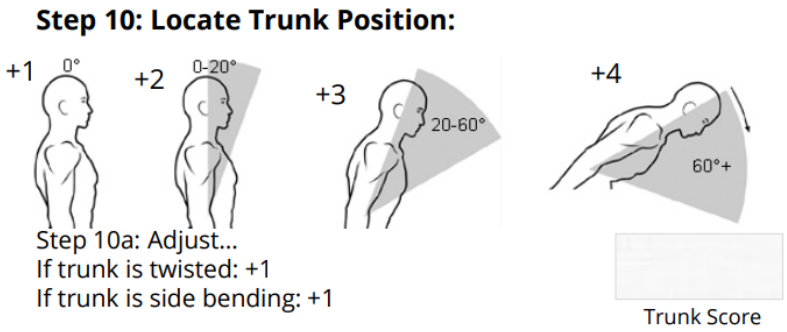

Abbildung 46 | Rumpf-Score [18]

Wie in Abbildung 47 zu sehen ist, werden die Beine bei der RULA-Methode nicht genau betrachtet. Sie unterscheidet lediglich, ob die Beine unterstützt werden oder nicht. Das kann mithilfe der Posenschätzung aber nicht beurteilt werden. Deshalb wurden ähnlich zu den anderen Scores zwei Abstufungen für die Kniewinkel eingeführt. Damit kann überprüft werden, ob die Knie abgewinkelt sind oder nicht. Dafür wird der Winkel zwischen Ober- und Unterschenkel berechnet. Als Grenzwert wurde ein Kniewinkel von 120° definiert. Bei einem Kniewinkel größer als 120° wird der Bein-Score auf den Wert 1 gesetzt und darunter auf den Wert 2. Liegen für die beiden Beine unterschiedliche Ergebnisse vor, wird für die weitere Berechnung der höhere Score verwendet.

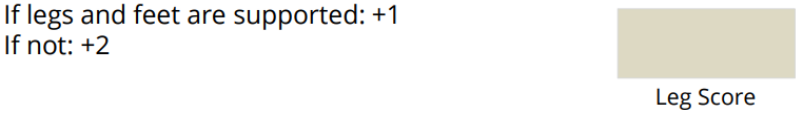

Abbildung 47 | Bein-Score [18]

Nacken-, Rumpf- und Bein-Score lassen sich nun auf einen Score mit dem Namen Nacken-Rumpf-Bein-Score zusammenfassen. Dazu wurde die Tabelle B aus dem RULA-Arbeitsblatt (hier in Tabelle 3) in Form einer Matrix in den Programmcode übertragen. Der Nacken-Score bestimmt dabei die benötigte Zeile der Matrix und die Kombination aus Rumpf- und Bein-Score ergibt die gesuchte Spalte.

Neck Posture Score	Table B: Trunk Posture Score											
	1		2		3		4		5		6	
	Legs		Legs		Legs		Legs		Legs		Legs	
	1	2	1	2	1	2	1	2	1	2	1	2
1	1	3	2	3	3	4	5	5	6	6	7	7
2	2	3	2	3	4	5	5	5	6	7	7	7
3	3	3	3	4	4	5	5	6	6	7	7	7
4	5	5	5	6	6	7	7	7	7	7	8	8
5	7	7	7	7	7	8	8	8	8	8	8	8
6	8	8	8	8	8	8	8	9	9	9	9	9

Tabelle 3 | Tabelle B im RULA-Arbeitsblatt zur Ermittlung des Nacken-Rumpf-Bein-Scores [18]

4.3.3 RULA-Score

Durch Kombination des Arm-Scores mit dem Nacken-Rumpf-Bein-Score lässt sich nun ein finaler RULA-Score bilden. Dieser wird über die Tabelle C im RULA-Arbeitsblatt (hier Tabelle 4) ermittelt, welche wieder in Matrixform in den Programmcode übertragen wurde. Der Arm-Score definiert dabei die Zeile und der Nacken-Rumpf-Bein-Score die Spalte. Dadurch erhält man einen Score zwischen eins und sieben, welcher die Ergonomie der aktuellen Pose beschreibt. Die Werte bedeuten dabei folgendes [18]:

- 1-2: vernachlässigbares Risiko, akzeptable Haltung
- 3-4: geringes Risiko, Änderung kann erforderlich sein
- 5-6: mittleres Risiko, weitere Untersuchung, bald ändern
- 7: sehr hohes Risiko, Veränderungen untersuchen und schnell umsetzen

Die ursprüngliche RULA-Methode beurteilt zusätzlich zur Körperhaltung auch die Muskelnutzung und die zu handhabenden Lasten. Bei sehr statischen oder oft wiederholenden Tätigkeiten sowie bei zu tragenden Lasten erhöht sich der ermittelte Score. Da dies aber nicht mit der Posenschätzung festgestellt werden kann, wird darauf nicht näher eingegangen.

Table C		Neck, Trunk, Leg Score						
		1	2	3	4	5	6	7+
Wrist / Arm Score	1	1	2	3	3	4	5	5
	2	2	2	3	4	4	5	5
	3	3	3	3	4	4	5	6
	4	3	3	3	4	5	6	6
	5	4	4	4	5	6	7	7
	6	4	4	5	6	6	7	7
	7	5	5	6	6	7	7	7
	8+	5	5	6	7	7	7	7

Tabelle 4 | Tabelle C im RULA-Arbeitsblatt zur Ermittlung des gesamten RULA-Scores [18]

4.4 Darstellung in Heatmaps

Für die Visualisierung der Ergebnisse wurden Heatmaps ausgewählt. Bei einer Heatmap handelt es sich um ein Diagramm, mit dem zweidimensionale Daten wie Punkte unter der Verwendung von Farben dargestellt werden können. Heatmaps dienen dazu, einen sehr einfachen und intuitiven Überblick über große Datenmengen zu geben. Der Name Heatmap lässt sich durch die normalerweise verwendeten Farben einer Wärmebildkamera herleiten. Diese sind für die meisten Menschen bekannt und leicht verständlich. Die Farbe Rot steht dabei für heiß, viel bzw. eine hohe Intensität. Bei den weiteren Abstufungen handelt es sich um Orange, Gelb, Grün und Blau. Letztere Farbe wird als kalt bzw. wenig interpretiert [51]. Verwendung finden Heatmaps beispielsweise in der Analyse des Konsumentenverhaltens in Supermärkten [9] oder Webseiten [10].

Die in dieser Diplomarbeit genutzten Heatmaps wurden halb transparent ausgeführt und enthalten im Hintergrund ein Bild des analysierten Arbeitsplatzes. Das Hintergrundbild zeigt dabei keine Menschen, sondern nur den Arbeitsbereich, der am aufgezeichneten Video zu sehen ist. Um dies zu erreichen, wird das Hintergrundbild aus dem Median der einzelnen Frames des Videos berechnet. Der Median ist dabei der Wert, der genau in der Mitte einer Datenverteilung liegt. Auf diese Weise können bewegte Elemente aus Bilderreihen herausgefiltert werden. Ein Bild besteht aus einer bestimmten Anzahl an Pixeln. Jedes Pixel enthält jeweils einen Wert für die Intensität der Farben Rot, Grün und Blau. Aus der Kombination dieser drei Farben lässt sich jede beliebige Farbe erzeugen. Bildet man nun den Median eines Farbwerts eines bestimmten Pixels von allen Bildern, so ergibt sich als Ergebnis der Wert, der am häufigsten auftritt. Führt man diesen Vorgang für die drei Farbwerte aller Pixel durch, erhält man als Ergebnis ein Bild, in dem jedes Pixel seine am häufigsten vorkommende Farbe enthält. Da bewegte Elemente nur eine kurzfristige Änderung der Farbe eines Pixels hervorrufen, werden sie durch die Medianbildung über alle Bilder des Videos herausgefiltert. Da sich die einzelnen Bilder des Videos nur geringfügig voneinander unterscheiden, wird hier nur jeder 15. Frame für die Medianberechnung verwendet. Dadurch kann die Rechenzeit verkürzt werden, ohne wichtige Informationen zu verlieren.

Die Darstellung der Ergebnisse aus der Ergonomieanalyse erfolgt mit drei verschiedenen Arten von Heatmaps. Zuerst wird eine Heatmap erzeugt, welche die Häufigkeit der Positionen eines Keypoints der 2D-Pose aufzeigt. Je häufiger der Keypoint an einer Stelle vorkommt, desto roter ist die Heatmap dort eingefärbt. Der zu untersuchende Keypoint kann vorher ausgewählt werden. Diese Heatmap enthält keine Informationen zur Ergonomie, sondern nur die Positionsdaten der Arbeitskraft. Beispielsweise lässt sich so, unter Verwendung des Keypoints des rechten Handgelenks, der Weg der rechten Hand darstellen. Dadurch können häufig gegriffene Gegenstände und eventuell zu lange Wege aufgezeigt werden. Durch die Analyse eines Fuß-Keypoints besteht die Möglichkeit, die Laufwege der Person abzubilden und gegebenenfalls zu verbessern. Im Programmcode werden die X- und Y-Koordinaten des ausgewählten Keypoints aus der Liste der 2D-Posen extrahiert und jeweils in einer eigenen Liste gespeichert. Um daraus eine Heatmap zu erhalten, wird ein Gitter mit den Dimensionen des Bildes erzeugt und jeweils untersucht, wie viele Punkte innerhalb einer 5x5-Pixelzelle liegen. Je nach Anzahl der eingeschlossenen Punkte ergibt sich dann die Farbe dieser Zelle. Das Ergebnis wird anschließend mit einer Gaußschen Kerndichteschätzung geglättet. Dadurch können Spitzen abgeschwächt und ein gleichmäßiger Verlauf erzielt werden. Die erhaltene Matrix wird anschließend als Heatmap über das Hintergrundbild geplottet. Abbildung 48 zeigt beispielhaft die Heatmap für einen Fuß-Keypoint.

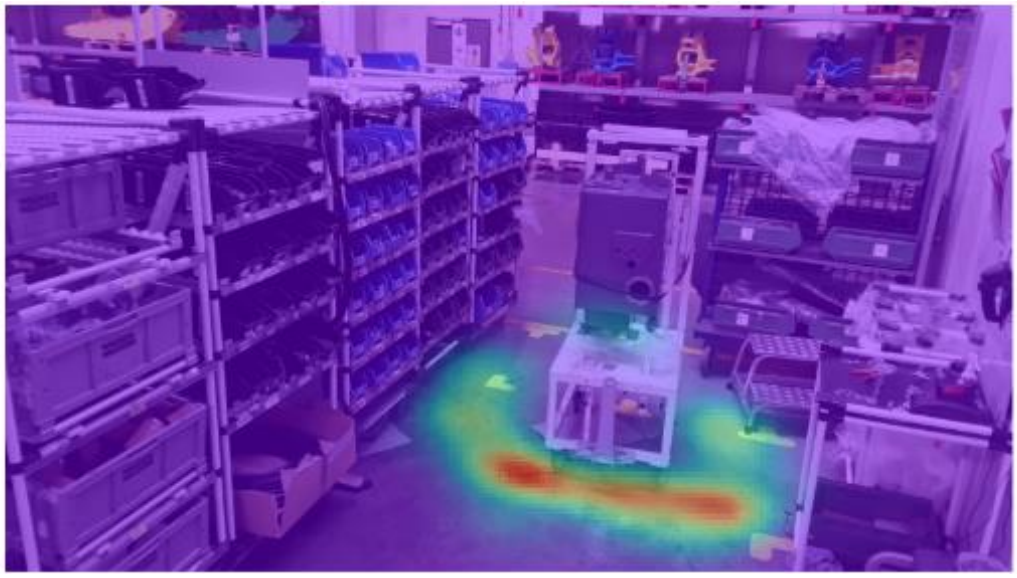

Abbildung 48 | Heatmap für die Häufigkeit der Positionen des Fuß-Keypoints

Die zweite erzeugte Heatmap stellt den berechneten RULA-Score bei der jeweiligen Position der Arbeitskraft dar. Liegt dabei an einer Stelle ein hoher RULA-Score vor, ist die Heatmap dort entsprechend roter eingefärbt. Dies ermöglicht es, Arbeitsstationen mit schlechter Ergonomie aufzuzeigen und die genaue Position der schädlichen Körperhaltung zu lokalisieren. Dadurch können gezielt Maßnahmen zur Verbesserung der Arbeitsbedingungen entwickelt werden. Um diese Heatmap zu erreichen, wird eine Matrix in der Größe des Bilds erzeugt. Jedes Element der Matrix steht dabei für ein Pixel des Bilds. Der Wert des Matrixelements bestimmt die Farbe des Pixels in der Heatmap. Beträgt der Wert 0, ergibt sich die Farbe Blau. Ein Wert von 7 wird rot dargestellt. Die RULA-Scores werden an der zugehörigen Position der Arbeitskraft in die Matrix eingetragen. Die Position entspricht der Mitte der beiden Fuß-Keypoints aus der 2D-Pose. Der Mittelpunkt der Füße berechnet sich dabei aus dem Mittelwert der X- bzw. Y-Koordinaten des linken und rechten Fußes. Sind in der Matrix auf allen vorhandenen Positionen die jeweiligen RULA-Scores eingetragen, wird die Auflösung dieser Matrix vergrößert, da man ansonsten die einzeln eingefärbten Pixel nicht sehen würde. Dafür wird die Matrix in Untermatrizen mit einer Größe von 15 Zeilen x 15 Spalten eingeteilt. Innerhalb dieser Untermatrizen wird dann der Median aus den Werten berechnet, die ungleich null sind. Anschließend wird der Median in jedes Element der Untermatrix eingetragen. Dadurch ergeben sich in der großen Matrix Felder mit einer Größe von 50x50, die denselben Wert enthalten. Diese Felder sind dann in der Heatmap besser erkennbar. Um fließendere Übergänge zu erhalten, kommt ein Gaußscher Filter zum Einsatz. Die sich ergebende Matrix wird abschließend als Heatmap über das Hintergrundbild geplottet. In Abbildung 49 ist ein Beispiel dargestellt. Der Vorgang erfolgt zusätzlich zum gesamten RULA-Score auch für den Arm- und den Nacken-Rumpf-Bein-Score. Durch die beiden zusätzlichen Heatmaps lässt sich besser nachvollziehen, wodurch ein hoher Score hervorgerufen wird.

Abbildung 49 | Heatmap des Ergonomie-Scores (dargestellt als Median der Ergonomie-Werte)

Da in dieser Heatmap nicht erkennbar ist, wie lange bzw. wie oft die Arbeitskraft eine schlechte Körperhaltung einnimmt, wird noch eine weitere Heatmap erzeugt. Diese bildet die Häufigkeit der gefährlichen Posen ab. Als gefährliche Posen wurden dabei diejenigen Posen definiert, welchen ein Score größer gleich fünf zugewiesen ist. Um diese Heatmap zu erreichen, werden abermals die X- und Y-Koordinaten des Mittelpunkts der Füße berechnet und nur dann in einer Liste gespeichert, wenn der Score der zugehörigen Pose fünf oder mehr beträgt. Anschließend wird wieder ein Gitter mit 5x5 Pixel großen Zellen erstellt und überprüft, wie viele der gespeicherten Punkte in der jeweiligen Zelle liegen. Je höher die Anzahl der Punkte in einer Zelle ist, desto roter wird sie dann dargestellt. Das bedeutet, dass an einer roten Stelle der Heatmap sehr viele gefährliche Posen vorliegen. Die Arbeitskraft verharrt dort also über eine längere Dauer in einer schlechten Körperhaltung. Um einen gleichmäßigeren Farbverlauf in der Heatmap zu erreichen, erfolgt abschließend eine Glättung mithilfe der Gaußschen Kerndichteschätzung. Das Ergebnis wird wiederum über das Hintergrundbild geplottet. Ein Beispiel ist in Abbildung 50 ersichtlich. Auch hier werden, zusätzlich zur Heatmap für den gesamten RULA-Score, Heatmaps für den Arm- und den Nacken-Rumpf-Bein-Score erzeugt. Dadurch lässt sich erkennen, welcher Faktor für den hohen Gesamt-Score verantwortlich ist.

Abbildung 50 | Heatmap für die Häufigkeit der gefährlichen Posen

4.5 Tracking einer Arbeitskraft

In den Videos können unter Umständen neben der Arbeitskraft auch unbeteiligte Personen auftreten. Der Posenschätzer kann allerdings nicht zwischen einzelnen Personen unterscheiden, sondern gibt alle erkannten Posen als Ergebnis aus. Die Reihenfolge in der Ausgabe hängt von der Position im analysierten Bild ab. Die am weitesten links gelegene Pose wird als erstes ausgegeben und die rechteste als letztes. Die erstellten Heatmaps enthalten die Daten von allen erkannten Posen. Läuft im Video eine Person durch das Bild, muss diese identifiziert und herausgefiltert werden, um die Heatmap der Arbeitskraft nicht zu verfälschen. Dafür wurde ein Tracking-Algorithmus implementiert, der auf dem Objekt-Tracker von Adrian Rosebrock basiert [52]. In dem Objekt-Tracker werden Gesichter erkannt und der geometrische Schwerpunkt des erzeugten Begrenzungsrahmens verfolgt. Der Algorithmus in dieser Diplomarbeit trackt die gesamte 2D-Pose.

Der entwickelte Tracking-Algorithmus basiert auf der Berechnung der Entfernung zwischen den erkannten 2D-Posen in aufeinanderfolgenden Frames. Im ersten Frame des Videos wird jeder erkannten 2D-Pose eine Identifikationsnummer (ID) zugeordnet. Die Posen werden dann gemeinsam mit ihrer ID abgespeichert. Beim darauffolgenden Frame wird versucht, die neu erkannten 2D-Posen mit den bereits bekannten zu verknüpfen. Dazu wird paarweise die Entfernung zwischen allen bekannten und allen neu erkannten Posen berechnet. Die Entfernung zwischen zwei Posen berechnet sich dabei aus der Addition aller Abstände zwischen gleichwertigen Keypoints. Sie ergibt sich also aus der Entfernung zwischen den Kopf-Keypoints plus der Entfernung zwischen den Nacken-Keypoints plus der Entfernung zwischen den linke-Schulter-Keypoints usw. Liegen alle Abstände zwischen den bekannten und den neu erkannten Posen vor, wird jeweils die Pose mit der geringsten Distanz zur bekannten Pose der

ID zugeordnet. Dies beruht auf der Annahme, dass sich eine Person zwischen aufeinanderfolgenden Frames bewegt, aber der Abstand zwischen ihren Posen kleiner ist, als zu allen anderen Posen. Wird im aktuellen Frame eine Pose mehr erkannt als bereits bekannt sind, so wird eine neue ID erstellt und gemeinsam mit der nicht zuweisbaren Pose gespeichert. Die beschriebenen Vorgänge werden dann für alle Frames des Videos wiederholt. Kann eine gespeicherte Pose im nachfolgenden Frame keiner neuen Pose mehr zugeordnet werden, weil etwa die Person das Bild verlassen hat, wird sie wieder gelöscht.

Als Ergebnis liefert der Tracking-Algorithmus die nach ID sortierten Posen. Dadurch besteht die Möglichkeit, einzelne IDs und somit einzelne Personen herauszufiltern. Am Anfang des Programmcodes können die IDs der Personen angegeben werden, die in der Analyse berücksichtigt werden sollen. Arbeiten an einem Arbeitsplatz mehrere Personen, ist es daher möglich, sowohl für jede einzelne Arbeitskraft als auch für alle gemeinsam die Heatmaps zu erzeugen. Unbeteiligte Personen, welche im Laufe des Videos durch das Bild gehen, werden vom restlichen Algorithmus dann ignoriert. Abbildung 51 zeigt ein erfolgreiches Beispiel des Trackings. Im linken Bild wurde die unbeteiligte Person erkannt. Durch den Tracking-Algorithmus wurde sie ignoriert, wie im rechten Bild zu sehen ist.

Abbildung 51 | Erfolgreiches Tracking (oben: zwei erkannte Mitarbeiter; unten: Mitarbeiter links wurde durch den Tracking-Algorithmus weggefiltert)

5 Durchführung des Feldversuchs

Während des Programmiervorgangs erfolgten die ersten Tests mit Beispielbildern aus dem Internet und Youtube-Videos. Aus den Videos wurden jeweils kurze Ausschnitte mit einer Länge von wenigen Sekunden erstellt, um die Wartezeit beim Ausführen des Codes zu verkürzen. Zusätzlich wurde ein Video selbst aufgezeichnet, um zu untersuchen, wie der Algorithmus Überkopfarbeit und Bücken beurteilt. Nach der Durchführung einiger Anpassungen erfolgte ein Feldversuch an zwei realen Montagearbeitsplätzen der Firma Wacker Neuson Linz GmbH.

Für die Auswahl geeigneter Arbeitsplätze wurden mehrere Kriterien definiert. Die zu filmende Tätigkeit sollte auf mehrere Stationen aufgeteilt sein, damit in der berechneten Heatmap auch unterschiedliche Bereiche markiert werden. Bei einer einzigen Station würden alle Posen an derselben Stelle liegen und somit wäre die Heatmap nur wenig aussagekräftig. Daher sollte der Arbeitsplatz zum Beispiel in Teilebereitstellung, Montagefläche und ähnlichem aufgeteilt sein. Des Weiteren muss der gesamte Arbeitsraum aus einer Position einsehbar sein, um von der Kamera erfasst werden zu können. Zusätzlich darf die Arbeitskraft von keinen anderen Objekten verdeckt werden, da dort sonst keine Posenschätzung möglich wäre. Um darüber hinaus auch demonstrative Heatmaps zu erzielen, sollten an dem Arbeitsplatz außerdem längere Laufwege zwischen den Stationen und eine schlechtere Ergonomie (Bücken, Überkopfarbeit) vorliegen.

In der Produktion der Wacker Neuson Linz GmbH wurden zwei Arbeitsplätze ausgewählt, die den Anforderungen entsprachen. Bei beiden handelt es sich um Vormontageplätze. Tätigkeiten direkt an der Montagelinie eigneten sich nicht zur Analyse, da dort sehr große Produkte wie Bagger und Dumper hergestellt werden und sich die Arbeitskräfte oft rund um das Erzeugnis bewegen. Somit wären sie für die Kamera häufig nicht sichtbar. Bei beiden Stationen wurde jeweils ein Arbeitstakt aufgenommen. Als Aufnahmegerät kam ein Samsung Galaxy Tab A zum Einsatz. Das Tablet wurde von der Firma inklusive einer geeigneten Halterung zur Verfügung gestellt. Mit der eingebauten Kamera lassen sich Videos mit einer Full-HD-Auflösung (1.920 x 1.080 Pixel) bei 30 Frames pro Sekunde aufzeichnen. Die hohe Auflösung führt zwar zu einer längeren Rechenzeit bei der Posenschätzung, wurde aber ausgewählt, um die Arbeitskraft auch bei größeren Entfernungen besser erfassen zu können. Die Halterung besitzt eine Schraubzwinge und lässt sich somit auf allen Vorrichtungen und Regalen mit rechteckigem Querschnitt sehr einfach befestigen. Bei den beiden Feldversuchen erfolgte die Montage der Kamera auf verschiebbaren Tafeln (wie in Abbildung 52), welche fixierbare Rollen besitzen. Diese wurden so positioniert und fixiert, dass die Kamera den gesamten Arbeitsraum erfasst. Die Fixierung der Tafel ist wichtig, um unabsichtliche Stöße oder Bewegungen zu

verhindern, welche die aufgezeichneten Bilder und somit auch die Keypoint-Koordinaten verändern würden.

Abbildung 52 | Befestigung des Tablets zur Erstellung des Videos

Das aufgezeichnete Video des Arbeitstaktes wurde in Google Drive hochgeladen, damit der Algorithmus darauf zugreifen kann. Die Posenschätzung musste auf mehrere Schritte aufgeteilt werden, da eine Analyse des ganzen Videos die Laufzeitbeschränkungen von Goolge Colab überschritt. Die Analyse der einzelnen Teile dauerte dann jeweils drei bis vier Stunden. Dabei wurde nur jeder 15. Frame des Videos analysiert. Dies entspricht einer Framerate von 2 Bildern in der Sekunde. Die Daten der 2D- und 3D-Posen wurden zwischenzeitlich in Google Drive gespeichert und anschließend für die Berechnung des Ergonomie-Scores und die Erzeugung der Heatmaps zusammengefügt.

Bei dem Großteil der Frames funktionierte die Schätzung der Pose, abgesehen von kleineren Ungenauigkeiten einzelner Gliedmaßen, sehr gut. Sogar weitgehend verdeckte Körperteile wurden erfolgreich geschätzt. Auch unbeteiligte Personen wurden durch den Tracking-Algorithmus erfolgreich ignoriert. Bei einer geringen Anzahl von Frames verlief die Posenschätzung jedoch nicht fehlerfrei. Dabei wurden Posen erkannt, obwohl sich an dieser Stelle kein Mensch befand (Abbildung 53 links und Mitte). Diese Fehler stellten aber kein Problem dar, weil durch den Tracking-Algorithmus diese zusätzlichen Posen in den weiteren Rechenschritten ignoriert wurden. Manchmal wurden für dieselbe Person zwei Posen ausgegeben, welche sich an derselben Stelle befanden und sich nur geringfügig voneinander unterschieden (Abbildung 53 rechts). Diese Fehler führten allerdings dazu, dass das Tracking fehl schlug. In diesem Fall aktualisierte der Tracker die Position der Arbeitskraft mit der zusätzlichen Pose und verlor anschließend die Person, wenn die falsche Pose wieder

verschwand. Diese Fehler mussten händisch ausgebessert werden, indem die hier zusätzlich erzeugte Personen-ID ebenfalls für das Tracking angegeben wurde. Das hatte den gleichen Effekt wie das Tracken von mehreren Personen. Dasselbe Problem entstand, wenn sich die Wege von zwei Personen kreuzten. Befanden sich die beiden Arbeitskräfte direkt hintereinander, war es auch möglich, dass der Tracking-Algorithmus die IDs der Personen vertauschte.

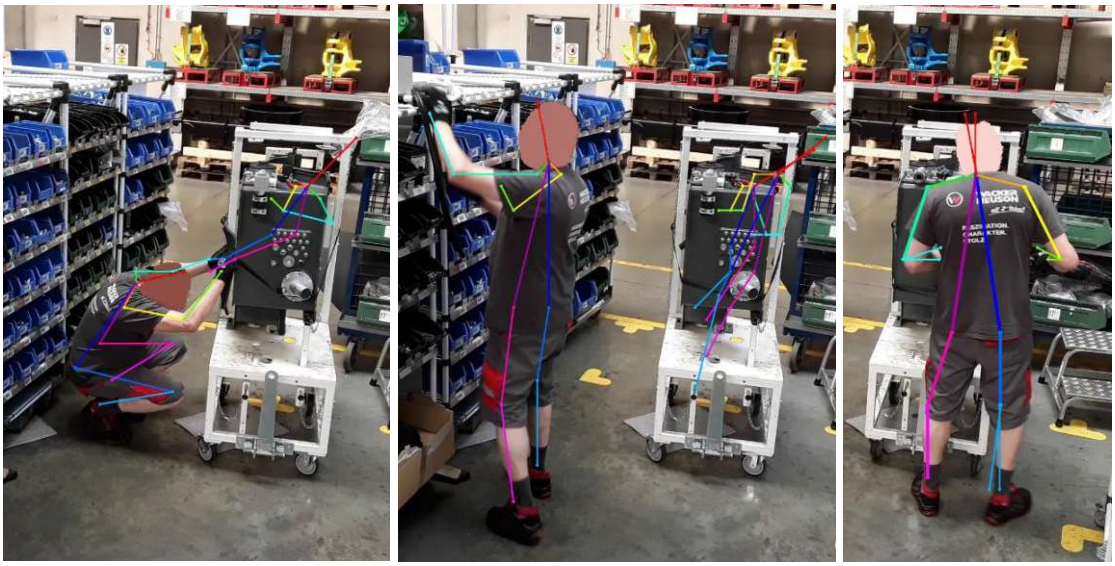

Abbildung 53 | Fehler des Posenschätzers

5.1 Tank-Vormontage

Beim ersten Arbeitsplatz für den Feldversuch handelte es sich um die Vormontage des Tanks eines Baggers. Abbildung 54 zeigt die analysierte Station, an welcher der bereits lackierte Grundkörper von einem Arbeiter mit zusätzlichen Teilen und Verschraubungen ausgestattet wird. Die dafür benötigten Teile werden in Sichtlagerboxen zur Verfügung gestellt (im Bild auf der linken Seite und rechts hinten). Das Werkzeug befindet sich auf einem Tisch (am rechten Bildrand) und auf der Wand, auf der die Kamera befestigt ist (am Bild nicht zu sehen). Der Montagevorgang dauert rund 20 Minuten. Der Tank befindet sich dabei auf einer fahrbaren Vorrichtung, welche für den Transport genutzt wird. Während des Arbeitsvorgangs ändert sich die Position des Tanks nicht.

Der Arbeitsplatz wurde ausgewählt, weil sich der Mitarbeiter hier während seiner Tätigkeit oft bücken und hinknien muss. Dadurch ergeben sich hohe Ergonomie-Scores, die auf den erzeugten Heatmaps erkennbar sind. Des Weiteren muss der Arbeiter beim Montagevorgang unterschiedliche Positionen einnehmen und fast der gesamte Arbeitsbereich kann von der Kamera erfasst werden. Einzig die Werkzeugbereitstellung, an der die Kamera befestigt wurde, ist nicht erkennbar. Da der Arbeiter dort aber ohnehin keine schädliche Haltung einnimmt, stellt dies keine besondere Einschränkung dar.

Abbildung 54 | Arbeitsplatz der Tankvormontage

5.2 Stoßstangen-Vormontage

Als zweiten Arbeitsplatz wurde die Vormontage der Stoßstange eines Dumpers ausgewählt. An dieser, in Abbildung 55 dargestellten, Station erfolgt die Vorbereitung der Stoßstange für die anschließende Montage auf der Linie. In dem Arbeitsvorgang werden verschiedene Teile, wie z.B. die Anhängekupplung und das Lüftungsgitter, an dem Stoßstangen-Rahmen angebracht. Die Rahmen werden in den fahrbaren Vorrichtungen (am rechten Bildrand) angeliefert. Die benötigten kleineren Teile befinden sich in Sichtlagerboxen, welche auf der linken und auf der hinteren Seite zu sehen sind. Größere Teile werden in dem Regal rechts hinten zur Verfügung gestellt. Das Werkzeug befindet sich auf der Wand und auf dem Tisch in der Mitte. Die zu montierenden Teile werden auf dem Tisch in der Mitte vorbereitet und dann an dem Stoßstangen-Rahmen angeschraubt. Die Montage erfolgt anfangs direkt in der Transportvorrichtung, indem sich der Arbeiter in den Rahmen beugt. Für die letzten Arbeitsschritte wird der Rahmen mit einem Hallenkran angehoben und in der Mitte der Arbeitsfläche platziert. Der gesamte Ablauf der aufgezeichneten Tätigkeiten an dieser Station dauert rund 15 Minuten.

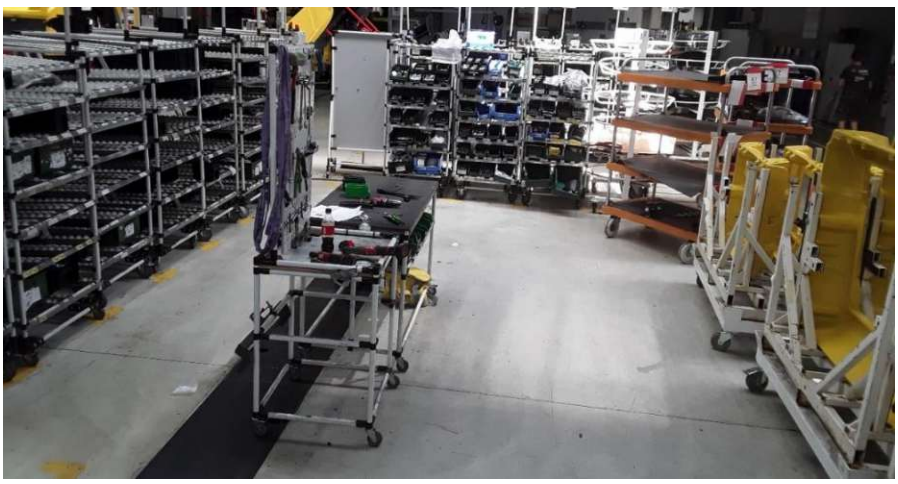

Abbildung 55 | Arbeitsplatz der Stoßstangenvormontage

Der Arbeitsplatz eignete sich für die Analyse, weil der Mitarbeiter hier längere Laufwege absolvieren muss und zusätzlich eine schlechte Körperhaltung bei der Montage einnimmt. Der gesamte Arbeitsbereich ist sehr gut einsehbar und die einzelnen Stationen liegen weiter auseinander. Dadurch können die Punkte in der Heatmap besser zum jeweiligen Arbeitsschritt zugeordnet werden. Abgesehen davon gehen hier manchmal unbeteiligte Personen durch den Arbeitsplatz, wodurch die Funktion des Tracking-Algorithmus sehr gut veranschaulicht werden kann.

6 Auswertung / Resultate

Dieses Kapitel beschreibt die Ergebnisse des durchgeführten Feldversuchs und beantwortet anschließend die Forschungsfragen aus der Einleitung.

6.1 Resultate des Feldversuchs

Im Rahmen des Feldversuchs wurden die, im vorigen Kapitel beschriebenen, Vormontage-Arbeitsplätze bei der Firma Wacker Neuson Linz GmbH für die Dauer von einem Takt gefilmt. Die aufgezeichneten Videos wurden mit dem erstellten Algorithmus analysiert. Dieser erzeugte verschiedene Heatmaps der Arbeitsstationen, welche die vorhandenen Ergonomieverhältnisse beschreiben. Im Folgenden werden die Ergebnisse des Feldversuchs vorgestellt.

6.1.1 Tank-Vormontage

Die erste erzeugte Heatmap soll die Laufwege des Arbeiters aufzeigen. Die Häufigkeit beider Fuß-Keypoints liefert dabei ein ähnliches Ergebnis. Deshalb wurde repräsentativ dafür die Häufigkeit der Positionen des Keypoints mit der Nummer 10 dargestellt. Beim Keypoint 10 handelt es sich um den rechten Fuß.

Abbildung 56 zeigt das Ergebnis der Analyse. Es ist gut ersichtlich, dass sich der Arbeiter rund um den Tank bewegt, wobei sich die häufigsten Arbeitsschritte auf der Vorderseite des Tanks befinden. Dies war auch eine neue Erkenntnis für die Führungskräfte der Wacker Neuson Linz GmbH. Sie nahmen bisher an, dass der Arbeiter auf allen Seiten gleichviele Tätigkeiten durchführt. Das meistgenutzte Material wird in denjenigen Sichtlagerboxen zur Verfügung gestellt, die sich in der Nähe des Tanks befinden. Die beiden Regale am linken Bildschirmrand werden hingegen selten bis gar nicht benötigt. Durch die kurze Distanz zur Teileandienung und zur Werkzeugbereitstellung auf der rechten Seite ergeben sich auch kurze Laufwege.

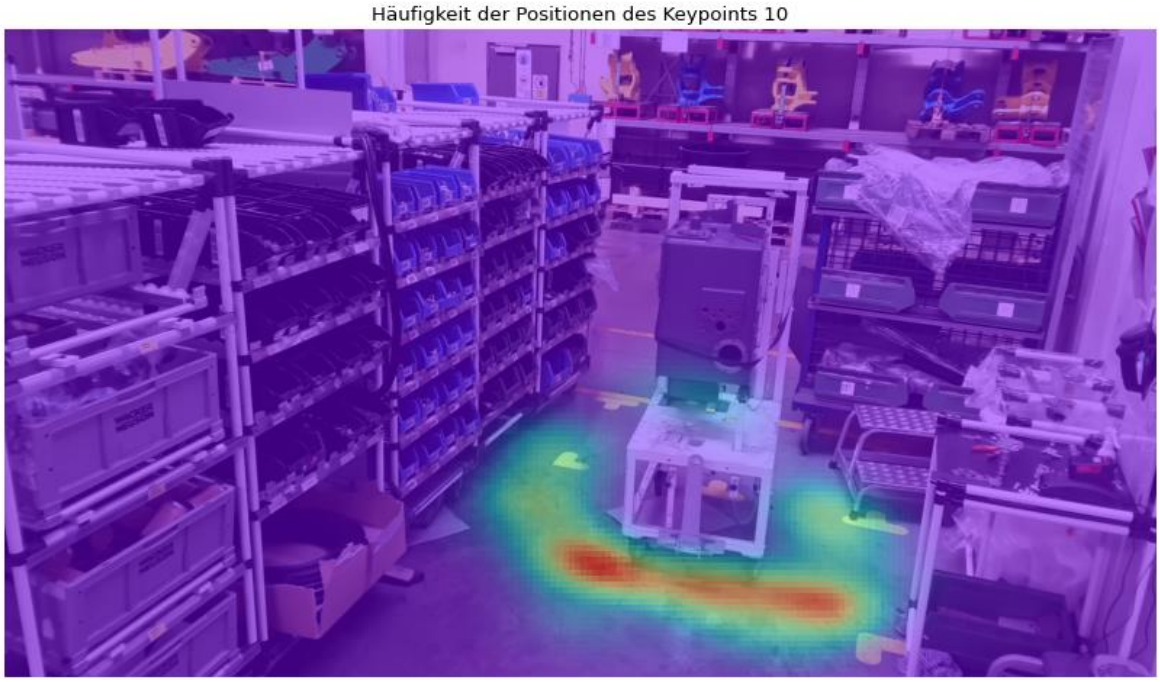

Abbildung 56 | Heatmap für die Darstellung der Laufwege bei der Tank-Vormontage

In Abbildung 57 wurden die Ergonomie-Scores bei der jeweiligen Position des Arbeiters aufgetragen. Zusätzlich zum Gesamt-Score wurden auch noch der Arm- und der Nacken-Rumpf-Bein-Score erzeugt, um beurteilen zu können, wodurch sich ein hoher Gesamt-Score zusammensetzt. Bei den eingezeichneten Punkten handelt es sich jeweils um den Mittelwert der beiden Fuß-Keypoints. Hier ist auch gut zu erkennen, dass sich der Mitarbeiter rund um den Tank bewegt. Sogar die Positionen hinter dem Tank wurden von dem Posenschätzer relativ gut geschätzt, obwohl der Arbeiter dort zum Großteil vom Tank verdeckt war. Des Weiteren sind in den Heatmaps auch offensichtlich fehlerhafte Punkte sichtbar, bei denen der Posenschätzer durch eine Überschneidung die Fuß-Keypoints falsch ermittelt hat. Dies ist zum Beispiel bei den beiden einzelnen Punkten in der oberen Tankhälfte der Fall.

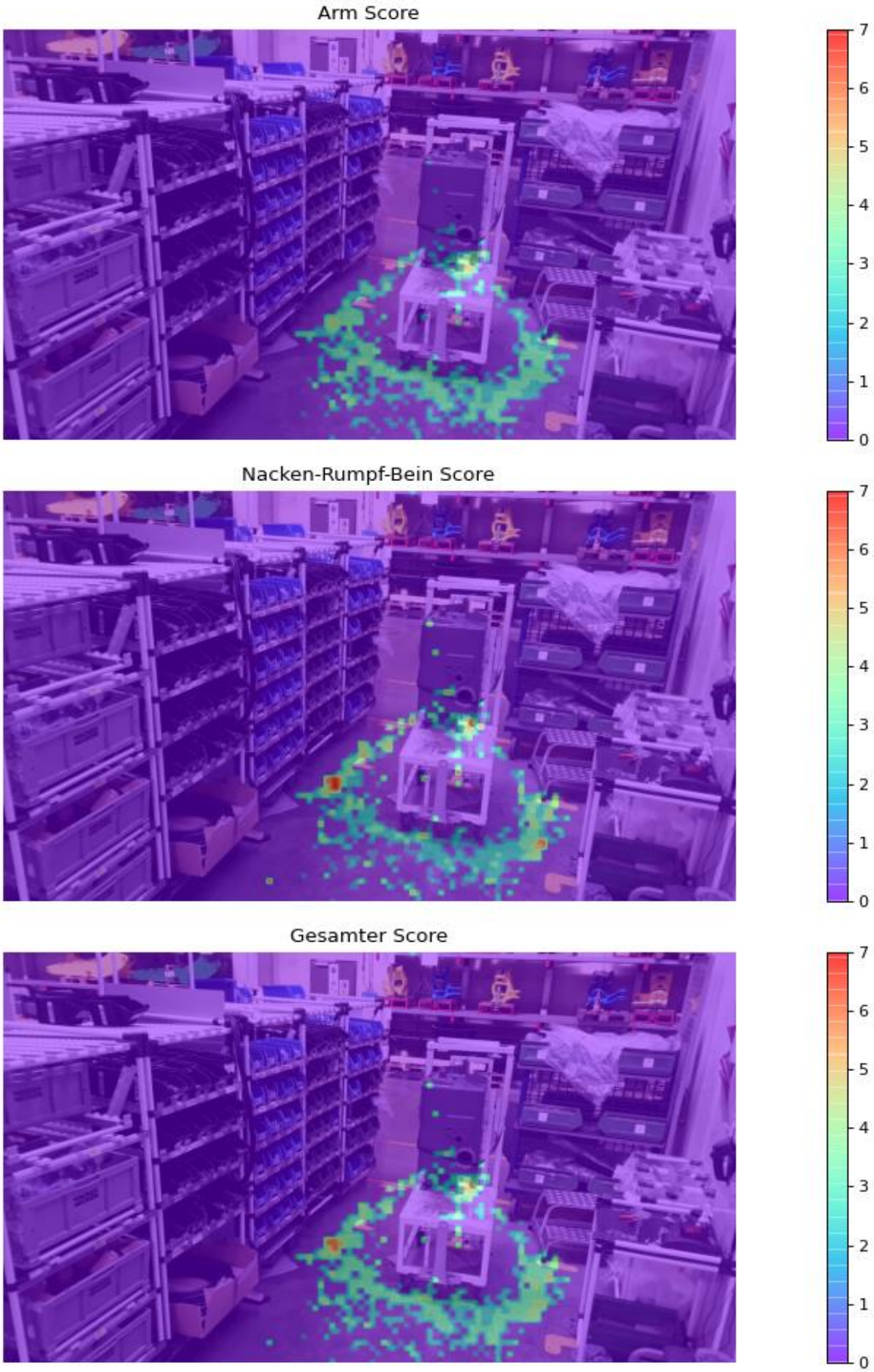

Abbildung 57 | Heatmaps der Ergonomie-Scores für die Tank-Vormontage

Bei Betrachtung der Heatmaps fällt sofort der rote Bereich links vom Tank auf. Dieser resultiert aus einer sehr schlechten Haltung von Nacken, Rumpf und Beinen. Tatsächlich muss sich der Arbeiter hier, wie in Abbildung 58 zu sehen ist, sehr oft bücken bzw. hinknien, um auf der Unterseite des Tanks Teile zu montieren. Auf der rechten Seite des Tanks befindet sich ein kleiner orangener Bereich in der Heatmap des Nacken-Rumpf-Bein-Scores. Dieser wird durch eine schlechte Nackenhaltung hervorgerufen, da der Mitarbeiter hier nach unten sieht und auf dem Tisch Schrauben und Beilagscheiben vorbereitet. Im Gesamt-Score ist dieser Bereich durch den gleichzeitig geringen Arm-Score nicht mehr sichtbar. Hinter dem Tank ist ein größerer gelber Bereich zu sehen. Dieser lässt sich dadurch erklären, dass dort der Arbeiter aus der obersten Etage des Regals auf der rechten Seite Teile entnimmt und damit oberhalb seiner Kopfhöhe greift. Außer den drei beschriebenen Bereichen sind keine besonderen Anhäufungen höherer Scores zu erkennen. An den meisten Stellen befindet sich der Score zwischen zwei und drei, welche noch als akzeptabel betrachtet werden können. Die RULA-Methode gilt allgemein als relativ konservative Methode, mit der Situationen öfter als risikobehaftet eingeschätzt werden als mit anderen Tools [53].

Abbildung 58 | Schlechte Körperhaltung an der rot markierten Stelle der Heatmap

Um nicht nur einen Überblick über die Lage und die Höhe der Ergonomie-Scores zu bekommen, wurden zusätzlich auch noch Heatmaps für die Häufigkeit der gefährlichen Posen berechnet. Dabei wurden nur die Positionen berücksichtigt, an denen ein Score größer gleich fünf vorlag. Durch diese Heatmaps kann besser eingeschätzt werden, wie lange der Arbeiter eine schlechte Körperhaltung einnimmt. Es macht natürlich einen Unterschied, ob die schädliche Pose für ein paar Sekunden oder für mehrere Minuten gehalten werden muss. Ein längeres Ausharren an derselben Position mit

dem Score von sieben würde in der Score-Heatmap in Abbildung 57 ebenso einen roten Punkt hervorrufen wie eine kurze Beanspruchung von einer Sekunde. Wie in Abbildung 59 zu sehen ist, tritt die schädliche Körperhaltung links vom Tank auch sehr häufig auf. Aufgrund dieser Erkenntnisse hat die Wacker Neuson Linz GmbH Maßnahmen eingeleitet, um das ständige Bücken und Knien des Arbeiters zu verringern. Die unergonomischen Tätigkeiten sollen dabei in Zukunft im Zuge eines späteren Arbeitsschritts durchgeführt werden, bei dem die Unterseite des Tanks besser erreichbar ist.

Abbildung 59 | Heatmap für die Häufigkeit der gefährlichen Posen für die Tank-Vormontage

6.1.2 Stoßstangen-Vormontage

Um die Laufwege bzw. häufigsten Positionen des Arbeiters bei der Stoßstangen-Vormontage darzustellen, wurde wiederum die Heatmap für die Häufigkeit des rechten Fußes berechnet. Abbildung 60 zeigt das Ergebnis dieser Analyse. Der Arbeiter hält sich dabei am häufigsten bei der Vorrichtung vom Stoßstangenrahmen auf. Dort trifft er einige Vorbereitungen wie z.B. Gewindebohren und bringt die Anhängekupplung an. Der grüne Bereich zwischen der Werkbank und der Stoßstangen-Vorrichtung resultiert aus dem späteren Anheben mit dem Hallenkran. Dabei hebt der Arbeiter die Stoßstange mit dem Kran auf die freie Fläche und führt dort die letzten Arbeitsschritte im aufgehangenen Zustand aus. Die Laufwege sind hier weiter, da zwischen der Werkbank und der Stoßstange ein paar Meter Abstand sind und der Arbeiter zum Werkzeugwechsel immer hin und her gehen muss. Für die Montage werden nur wenige Teile aus den Sichtlagerboxen benötigt, weshalb hier nur minimale Verfärbungen in der Heatmap erkennbar sind.

Abbildung 60 | Heatmap für die Darstellung der Laufwege bei der Stoßstangen-Vormontage

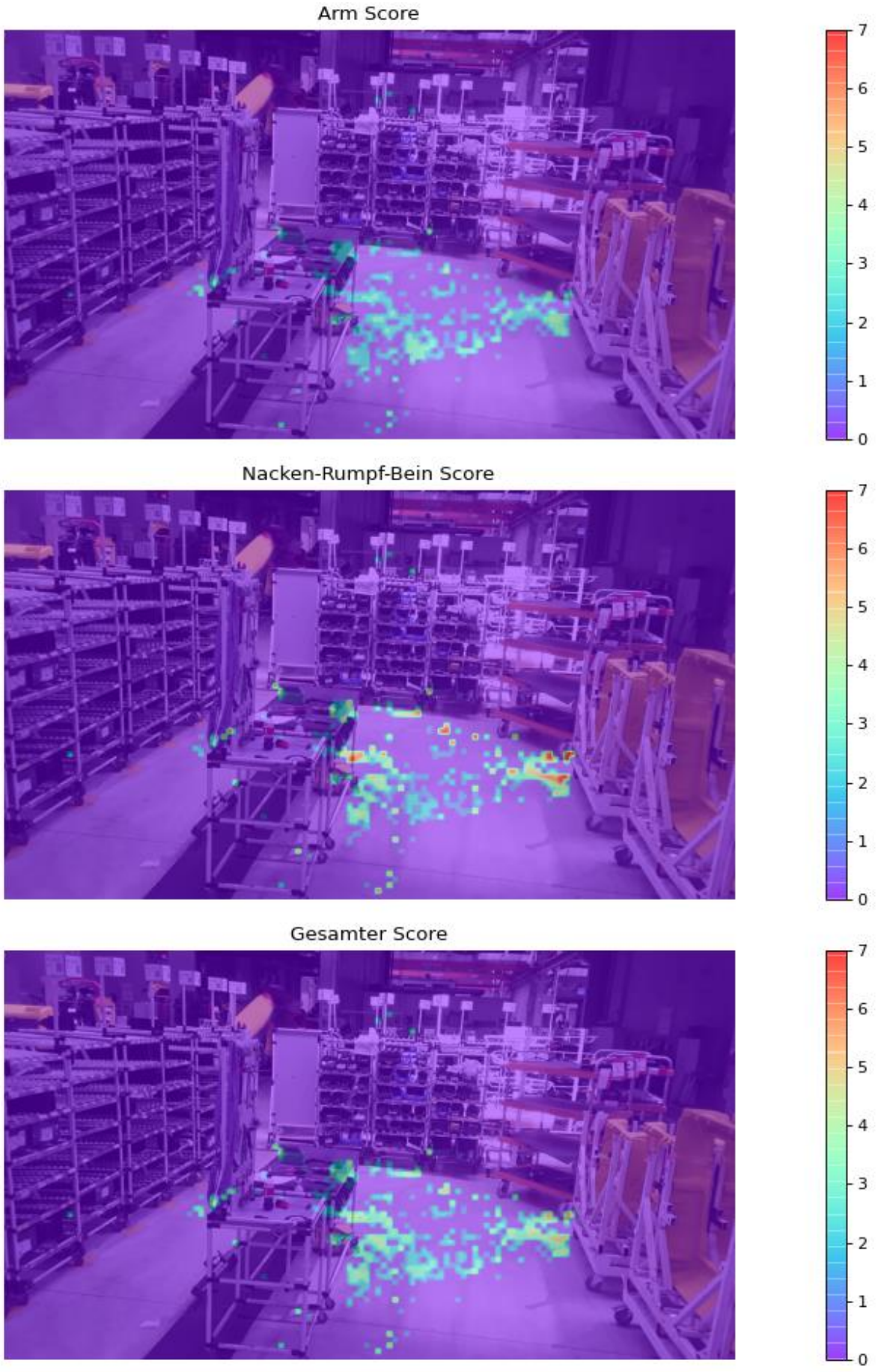

Abbildung 61 | Heatmaps der Ergonomie-Scores für die Stoßstangen-Vormontage

In Abbildung 61 ist wiederum der berechnete Ergonomie-Score auf der jeweiligen Position der Mitte der Fuß-Keypoints aufgetragen. Dabei sind auch wieder ein paar fehlerhafte Punkte erkennbar, wie z.B. unter dem Werkzeugtisch. Sehr gut zu sehen sind die Regale, aus welchen die Teile herausgenommen wurden. Auffällig bezüglich der Ergonomie ist die Position bei der Stoßstangenrahmen-Vorrichtung auf der rechten Seite des Bildes. Vor allem der Nacken-Rumpf-Bein-Score weist hier sehr hohe Werte auf. Der Gesamt-Score fällt dort jedoch geringer aus, weil der Arm-Score niedriger ist. Dies resultiert daraus, dass sich der Arbeiter hier sehr stark bückt und dabei, wie in Abbildung 62 ersichtlich ist, teilweise die Tätigkeiten im Rahmen durchführt. Dadurch sind die Arme für die Kamera manchmal nicht erkennbar. Die roten Punkte im Nacken-Rumpf-Bein-Score bei den Regalen folgen daraus, dass die benötigten Teile in der untersten Ebene bereitgestellt werden. Bei der weiteren rötlichen Verfärbung hinter der Werkbank montiert der Arbeiter das Wacker Neuson Logo auf das Lüftungsgitter. Durch die sehr kleinen Schrauben beugt sich der Arbeiter hier stark über den Tisch und erzeugt dadurch den hohen Score.

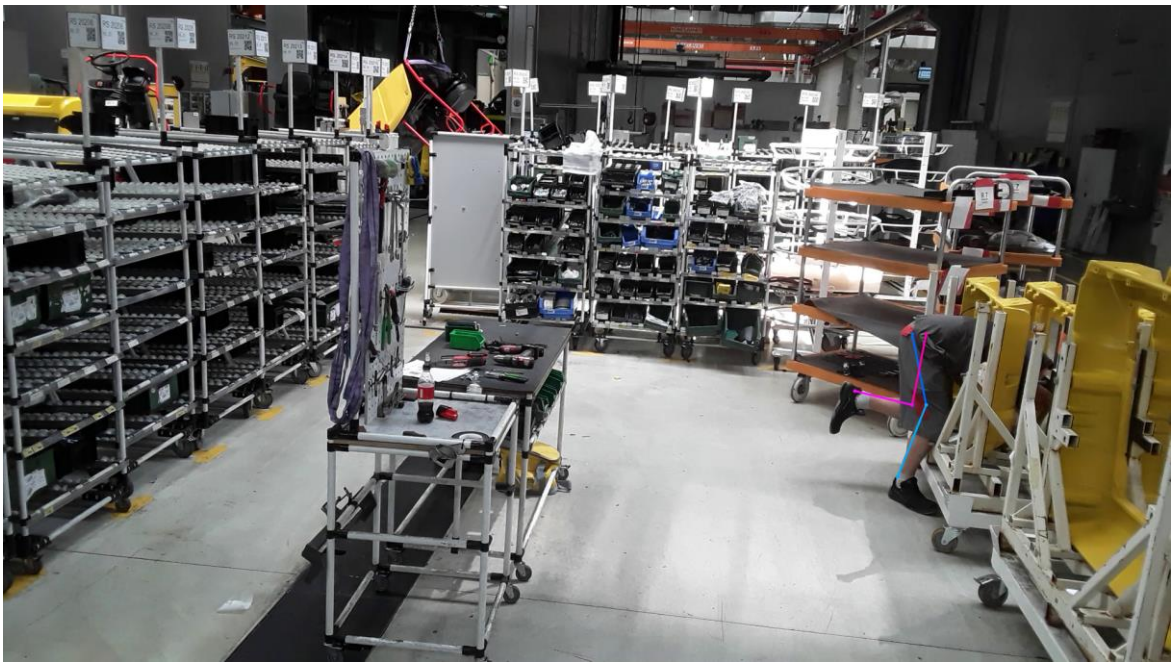

Abbildung 62 | Schlechte Körperhaltung sowie nicht sichtbare Arme bei der Stoßstangen-Vormontage

Abbildung 63 zeigt die Häufigkeit der gefährlichen Posen. Dabei ist ersichtlich, dass die Arme an dieser Arbeitsstation nur bei einer sehr geringen Anzahl an Frames eine schlechte Haltung einnehmen. Der restliche Körper befindet sich hingegen sehr häufig bei der Stoßstangen-Vorrichtung in einer gefährlichen Position. Auch bei diesem Arbeitsplatz hat die Wacker Neuson Linz GmbH aufgrund der hier erzielten Erkenntnisse Maßnahmen zur Verbesserung der Ergonomie eingeleitet. Um das ständige Bücken zu vermeiden, soll in Zukunft eine Vorrichtung zum Einsatz kommen, mithilfe der die Vormontage in einer ergonomischeren Arbeitshöhe durchgeführt werden kann.

Abbildung 63 | Heatmap für die Häufigkeit der gefährlichen Posen für die Stoßstangen-Vormontage

6.2 Resultate in Bezug auf die Forschungsfragen

Im Folgenden werden die Forschungsfragen aus der Einleitung beantwortet:

Welche Methoden zur Analyse der Ergonomie mit Kameras gibt es bereits?

Im Rahmen der Literaturrecherche wurden im Kapitel 3 aktuelle Algorithmen für die Posenschätzung sowie die Anwendung dieser auf die Ergonomieanalyse vorgestellt. Aus den erhaltenen Daten der Posenschätzer ermitteln die beschriebenen Methoden die Winkel zwischen den einzelnen Körperteilen. Diese Winkel werden anschließend mit existierenden Ergonomie-Beobachtungsmethoden bewertet. Die Ergebnisse werden dabei auf unterschiedliche Art und Weise dargestellt. [36] entscheidet beispielsweise nur, ob ein vorliegender Winkel gefährlich ist oder nicht. [37] ermittelt einen Index, der die Ergonomie des gesamten Arbeitsvorgangs beschreibt. [38] und [39] stellen die Ergebnisse in einem Diagramm dar, in dem Ergonomie-Scores über die Zeit aufgetragen sind.

Zusätzlich wurden auch Methoden aufgezeigt, die die Ergonomie ohne die Schätzung der menschlichen Pose bewerten. Dafür werden beispielsweise Inertial Measurement Units (IMUs) oder Tiefenkameras genutzt. Die IMUs liefern ebenfalls Informationen über die Gelenkwinkel, welche in [19] für die Ergonomieanalyse verwendet werden. [11] nutzt ein Netzwerk von Tiefenkameras, um eine Heatmap der Positionen einer Arbeitskraft zu erzeugen. Die Ergonomieverhältnisse eines Arbeitsplatzes wurden allerdings bisher noch nie in einer Heatmap dargestellt.

Wie kann der Mensch in einem zweidimensionalen Video erkannt werden?

Die Erkennung des Menschen in einem Video erfolgt mittels eines Posenschätzers. Die Posenschätzer basieren auf faltenden neuronalen Netzwerken und müssen mit geeigneten Datensätzen trainiert werden. Sie benötigen als Eingabe ein Bild und liefern als Ergebnis die 2D- bzw. 3D-Koordinaten der Keypoints der erkannten Personen. Das faltende neuronale Netzwerk untersucht dabei die einzelnen Pixel auf bestimmte Muster, um so z.B. einen Ellbogen zu erkennen. Der Algorithmus, welcher im Rahmen der Diplomarbeit erstellt wurde, verwendet den Posenschätzer aus [27]. Dieser verfolgt den sogenanntem 2D-to-3D-Lifting-Ansatz, welcher zuerst die 2D-Pose ermittelt und daraus eine 3D-Pose erzeugt. Das hat den Vorteil, dass sowohl die 2D- als auch die 3D-Pose ausgegeben und für die anschließende Analyse verwendet werden können.

Wie können die Bewegung einzelner Körperteile und die Ergonomie in Heatmaps dargestellt und interpretiert werden?

Um die Heatmaps zu erzeugen, werden die erhaltenen Daten aus der Posenschätzung verwendet. Die 2D-Daten repräsentieren dabei die Position der Person auf dem Bild und die 3D-Daten die Körperhaltung. Unter Verwendung der 2D-Koordinaten eines Keypoints kann die Häufigkeit der Positionen eines Körperteils in einer Heatmap dargestellt werden. Betrachtet man einen Fuß-Keypoint, werden die Laufwege und häufigsten Positionen der Arbeitskraft aufgezeigt. Mithilfe der 2D-Koordinaten einer Hand ist es möglich, häufig gegriffene Gegenstände zu identifizieren.

Um die Ergonomie an einem Arbeitsplatz zu bewerten, werden aus den erhaltenen 3D-Koordinaten die Winkel zwischen den einzelnen Körperteilen berechnet. Mit der RULA-Methode als Vorbild, werden anschließend aus den Gelenkwinkeln Ergonomie-Scores berechnet. Die RULA-Methode weist dabei den jeweiligen Gelenkwinkeln einen Score zu, welcher zwischen eins und sieben liegt und ein Maß für die Ergonomie darstellt. Dieser Ergonomie-Score kann dann an der jeweiligen Position der Arbeitskraft in der Heatmap eingetragen werden. Ein hoher Wert erzeugt dabei eine rote Farbe und ein niedriger eine blaue Farbe. Dadurch wird aufgezeigt, an welchen Stellen der Arbeitsstation eine schlechte Ergonomie vorliegt.

7 Diskussion und Ausblick

In dieser Diplomarbeit wurden aktuelle Algorithmen zur 2D- und 3D-Posenschätzung sowie die Anwendung dieser für die Bewertung der Ergonomie eines Arbeitsplatzes aufgezeigt. Die Schätzung der menschlichen Pose befindet sich gerade in der Forschung, weshalb laufend neuere und bessere Algorithmen veröffentlicht werden. Die Verbesserungen werden dabei durch verschiedene Methoden der Keypoint-Ermittlung, mehrstufige Netzwerke oder Tracking-Algorithmen erzielt. Letztere tragen dazu bei, die Pose von stark verdeckten Menschen zu schätzen. Des Weiteren ist es auch bereits möglich, Gesichtszüge oder kleinere Körperteile wie Finger zu erkennen.

Die dreidimensionale Posenschätzung aus einem zweidimensionalen Bild stellt eine große Herausforderung dar. Insbesondere die Ermittlung der Körperhaltung mehrerer Personen und deren Position im Raum ist immer noch schwierig, da sie hohe Anforderungen an den Speicher und die Leistung stellen. Eine Vielzahl von 3D-Posenschätzungsmethoden verfolgt den 2D-to-3D-Lifting-Ansatz, bei dem zuerst die 2D-Pose und anschließend daraus die 3D-Pose ermittelt wird. Ein solcher wurde auch in dieser Diplomarbeit genutzt. Der Vorteil des 2D-to-3D-Lifting-Ansatzes ist, dass dieser sowohl die 2D- als auch die 3D-Pose ausgibt. Die 2D-Pose wurde dabei für die Position der Arbeitskraft im Bild benötigt und die 3D-Pose für die Bewertung der Körperhaltung. Aus diesen Daten konnten dann Heatmaps erzeugt werden, welche die Ergonomie an dem jeweiligen Arbeitsplatz darstellen.

Im Gegensatz zu den existierenden Ergonomiebewertungsmethoden wurde als Darstellungsform der Ergebnisse die Heatmap ausgewählt. Bisherige Methoden stellten die Ergebnisse beispielsweise nur als Diagramm über die Zeit dar. Die Heatmaps ermöglichen aber auch eine räumliche Interpretation, wodurch sofort einzelne Arbeitsstationen mit einer schlechten Körperhaltung erkannt werden können.

7.1 Nutzen der Ansätze und Ergebnisse

Die aktuellen Entwicklungen bei künstlichen neuronalen Netzwerken für die Posenschätzung ermöglichen eine automatische Analyse der Arbeitsbedingungen mit geringem Kosten- und Personenaufwand. Abgesehen von einem leistungsstarken Rechner erfordert die Methode nur ein Stativ und eine handelsübliche Handykamera oder ähnliches. Der entwickelte Algorithmus erzeugt aus dem aufgezeichneten Video automatisch die Heatmaps. Ein Mensch muss nach Durchführung des Trackings nur überprüfen, ob für alle Frames die Posendaten vorliegen. Durch Fehler des Posenschätzers besteht nämlich die Möglichkeit, dass der Tracking-Algorithmus die Arbeitskraft verliert. Dies ist der Fall, wenn die Pose der Person falsch oder doppelt geschätzt wird. Ist die falsch erkannte Pose dann näher an der Pose aus dem vorigen Frame als die richtige Pose der Arbeitskraft, so verliert der Tracking-Algorithmus die

Person, wenn die irrtümlich erzeugte Pose im darauffolgenden Frame wieder verschwindet. In diesem Fall werden ab diesem Frame keine Posendaten mehr ausgegeben. Der Tracking-Algorithmus erzeugt aber für jede erkannte Pose, die keiner Pose aus dem vorigen Frame zugeordnet werden kann, eine neue ID. Deshalb muss bei den zu trackenden IDs nur diese zusätzliche ID händisch ergänzt werden. Dieser Vorgang erfordert ungefähr zehn Minuten Personenaufwand, welcher unter der Verwendung eines besseren Posenschätzers noch weiter reduziert werden kann. Die restliche Analyse erfolgt automatisch durch den Computer. Dadurch kann erheblicher Aufwand bei der Ergonomiebewertung eingespart werden, für die sonst oft eigene Ergonomie-Fachkräfte zum Einsatz kommen. Zusätzlich ist das Ergebnis nicht mehr von den subjektiven Wahrnehmungen der Fachleute abhängig. Es wird nur noch durch die Genauigkeit des Posenschätzers beeinflusst, welche sich von Netzwerk zu Netzwerk unterscheidet und sich in Zukunft mit Sicherheit noch weiter verbessert. Die Arbeitskraft kann ihre Tätigkeiten ohne Beeinträchtigungen (wie etwa durch einen IMU-Anzug) verrichten und wird unter Umständen von einer Kamera weniger abgelenkt als von einer beobachtenden Person. Außerdem ist die Arbeitskraft in den Ergebnissen nicht sichtbar, wodurch eine gewisse Anonymität sichergestellt wird. Durch den Tracking-Algorithmus besteht zusätzlich die Möglichkeit, mehrere Personen gleichzeitig zu analysieren und unbeteiligte Personen zu ignorieren.

Die Darstellung der Ergebnisse in Heatmaps bringt viele Vorteile. Sie fassen viele Informationen in einem Bild zusammen und liefern somit sehr schnell einen guten Überblick. Durch die allgemein bekannten Zusammenhänge der Farben Rot mit heiß bzw. viel und Blau mit kalt bzw. wenig werden Heatmaps von allen Menschen leicht verstanden. Die Darstellung der Ergebnisse in Heatmaps ermöglicht somit eine einfache Interpretation der Ergonomiebedingungen an einem spezifischen Arbeitsplatz. Ergonomisch schlechte Arbeitsplätze können dadurch auf einen Blick identifiziert und die Maßnahmen gezielter eingeleitet werden. Abgesehen von der Bewertung der Körperhaltung zeigt die erstellte Heatmap auch die Laufwege der Arbeitskraft auf. Unter Verwendung derselben Daten wurden zusätzlich auch noch weitere Heatmaps zur Darstellung der Häufigkeiten von bestimmten Keypoints und gefährlichen Posen erzeugt. Diese dienen dazu, den Arbeitsplatz noch besser zu verstehen. Die Häufigkeitsverteilungen einzelner Keypoints (wie z.B. der Fuß-Keypoints) zeigen oft aufgesuchte Positionen und genutzte Gegenstände auf. Dadurch können die Objekte optimaler platziert werden. Die Darstellung der Häufigkeit von gefährlichen Posen dient dazu, ein besseres Verständnis über die Dauer der schädlichen Körperhaltungen zu geben.

7.2 Einschränkungen der Ansätze und Ergebnisse

Neben den Vorteilen gibt es jedoch auch noch einige Einschränkungen bei der entwickelten Methode. Google Colab bietet zwar kostenlosen Zugriff auf GPUs ohne

notwendige Installationen oder Konfigurationen, jedoch gibt es bei der Nutzung auch ein paar Restriktionen. Die maximale Laufzeit der Notebooks beträgt zwölf Stunden und bei Inaktivität wird die Verbindung zu den Servern häufig getrennt, um die Ressourcen von Google anderen Personen zur Verfügung zu stellen [54]. Dadurch war es nicht möglich, den Programmcode über Nacht auszuführen und die Videos mussten für die Posenschätzung jeweils in drei kürzere Teile aufgeteilt werden. Nach der erfolgten Posenschätzung mussten die erhaltenen Daten dann wieder zusammengefügt werden. Die Schätzung der Pose war dabei der rechenaufwändigste Schritt der Analyse und dauerte für ein 15-minütiges Video ca. neun Stunden. Die Dauer der Berechnung hängt aber stark von der Auflösung des Videos ab. In diesem Versuch wurden die Videos mit einer Full-HD Auflösung erstellt. Bei einer geringeren Auflösung wäre die Rechenzeit deutlich geringer. Die hohe Auflösung wurde jedoch ausgewählt, um möglichst genaue Posen zu erhalten, selbst wenn die Arbeitskraft sehr weit entfernt ist. Durch die Verwendung von Google Colab könnten zusätzlich bei anwendenden Unternehmen Datenschutzbedenken aufkommen, da die zu analysierenden Videos auf die Server von Google Drive geladen werden müssten. Zur Aufzeichnung des Videos ist des Weiteren auch das Einverständnis von dem Betriebsrat der Firma und der zu filmenden Arbeitskraft notwendig.

Der zu analysierende Arbeitsplatz muss auch gewisse Anforderungen erfüllen. Der gesamte Arbeitsraum sollte dabei aus einer Position einsehbar sein, um die Arbeitskraft zu jedem Zeitpunkt aufzeichnen zu können. Zusätzlich darf der Person von keinen anderen Objekten verdeckt werden, da dort die Ergonomie sonst nicht beurteilt werden kann. Daher ist es eher schwierig, Arbeitsplätze mit sehr großen Produkten zu analysieren, wenn die Arbeitskraft auf allen Seiten des Produkts Tätigkeiten ausführt. Alternativ kann in solchen Fällen der Arbeitsvorgang von mehreren Seiten gefilmt werden, um dann Heatmaps aus den verschiedenen Blickwinkeln zu erhalten. Weiters sollte der Arbeitsplatz in mehrere Stationen aufgeteilt sein, sodass beispielsweise die Teilebereitstellung und die Montageflächen an unterschiedlichen Plätzen angeordnet sind. Ansonsten wäre eine Heatmap sinnlos, da sich dann alle Punkte an derselben Stelle befinden würden.

Das verwendete neuronale Netzwerk zur Posenschätzung erzeugt manchmal fehlerhafte Posen, die in Probleme bei dem Tracking-Algorithmus resultieren. Dadurch ist eine manuelle Nacharbeit notwendig. Die ermittelten Posen sind auch oft nicht ganz exakt. Die Keypoints weichen von den wirklichen Gelenkpunkten geringfügig ab, wodurch sich auch Ungenauigkeiten in der 3D-Pose ergeben. Diese sollten sich aber im Schnitt ausgleichen. Das wurde auch in den Feldversuchen bewiesen, welche die schädlichen Körperhaltungen aufzeigten.

Im Allgemeinen ist die RULA-Methode eine relativ konservative Methode, mit der Situationen öfter als risikobehaftet eingeschätzt werden als mit anderen Tools [53].

Dies ist auch in den Heatmaps der Ergonomie-Scores ersichtlich, in denen sehr häufig ein Score von 3 auftritt. Ein Wert von 3 bedeutet dabei bereits, dass weitere Untersuchungen und möglicherweise auch Veränderungen durchgeführt werden sollten. Die roten Flecken in den Heatmaps, welche für Scores von 7 stehen, repräsentieren aber auch in der Realität schädliche Körperhaltungen und wurden somit vom Algorithmus richtig erkannt.

Nicht alle Kriterien der RULA-Methode können mit der kamerabasierten Analyse beurteilt werden. Der Posenschätzer erkennt nur den Menschen auf dem Bild. Es ist daher nicht unterscheidbar, ob Arme bzw. Beine unterstützt werden oder ob sich die Person anlehnt. Kleinere Bewegungen wie z.B. das Anheben der Schulter oder grundsätzlich alle Bewegungen des Handgelenks können auch nicht erfasst werden. Das Arbeitsblatt der RULA-Methode berücksichtigt zusätzlich zur Körperhaltung auch die Muskelnutzung und die getragene Last. Diese Faktoren lassen sich mit einer Kamera ebenfalls nicht aufzeichnen.

Ein Nachteil der Heatmaps ist, dass sie nur Informationen über den Ort und keine über den Zeitpunkt und die Tätigkeit der schlechten Körperhaltung bieten. Führt die Arbeitskraft an einer Position verschiedene Tätigkeiten aus, muss also anschließend noch überprüft werden, welche Tätigkeit die schädliche Haltung hervorruft. Der zu analysierende Bereich wird aber stark eingeschränkt, da die Position bereits bekannt ist. Die Heatmaps bieten einen guten Überblick über eine Arbeitsstation und eignen sich daher sehr gut für eine erste Analyse und als Präsentationsmaterial.

7.3 Nächste mögliche Schritte zur Weiterentwicklung

Um die Ressourcenbeschränkungen von Google Colab und eventuelle Datenschutzbedenken zu vermeiden, besteht die Möglichkeit, eine Offline-Lösung zu realisieren. Der entwickelte Programmcode kann dabei ohne Änderungen verwendet werden. Dafür ist jedoch eine geeignete Hard- und Software notwendig. Der Computer sollte eine möglichst leistungsfähige Grafikkarte besitzen, um die hohe Anzahl an Rechenschritten in einer akzeptablen Zeit durchzuführen. Des Weiteren müssen eine Python-Programmierumgebung sowie die benötigten Module installiert werden. Unter diesen Voraussetzungen ist es dann möglich, das ganze Video in einem Teil zu analysieren und die Berechnung über Nacht auszuführen.

Es werden laufend neue und bessere Algorithmen zur Posenschätzung veröffentlicht. Daher besteht in Zukunft die Möglichkeit, einen anderen Posenschätzer in den Programmcode zu implementieren. Dies kann die Fehler des aktuell verwendeten Posenschätzers beheben, die zu den Verwechslungen beim Tracking führen. Ein genauerer Posenschätzer würde auch in einer Verbesserung der Posendaten resultieren. Dadurch würden die 3D-Posen besser mit der realen Körperhaltung übereinstimmen, wodurch sich auch die Qualität der Score-Heatmaps erhöht.

8 Literaturverzeichnis

[1] D. Wang, F. Dai und X. Ning, „Risk Assessment of Work-Related Musculoskeletal Disorders in Construction: State-of-the-Art Review," *Journal of Construction Engineering and Management,* Februar 2015.

[2] X. Li, S. Han, M. Gül, M. Al-Hussein und M. El-Rich, „3D Visualization-Based Ergonomic Risk Assessment and Work Modification Framework and Its Validation for a Lifting Task," *Journal of Construction Engineering and Management,* Juli 2017.

[3] A. Alwasel, K. Elrayes, E. M. Abdel-Rahman und C. Haas, „Sensing Construction Work-Related Musculoskeletal Disorders (WMSDs)," Juni 2011.

[4] J. Du und V. G. Duffy, „A methodology for assessing industrial workstations using optical motion capture integrated with digital human models," *Occupational Ergonomics,* Jänner 2007.

[5] S. C. Puthenveetil, C. P. Daphalapurkar, W. Zhu, M. C. Leu, X. F. Liu, J. K. Gilpin-Mcminn und S. D. Snodgrass, „Computer-automated ergonomic analysis based on motion capture and assembly simulation," *Virtual Reality,* Nr. 19, p. 119–128, 2015.

[6] S. Han und S. Lee, „A vision-based motion capture and recognition framework for behavior-based safety management," *Automation in Construction,* Nr. 35, p. 131–141, 2013.

[7] S. J. Ray und J. Teizer, „Real-time construction worker posture analysis for ergonomics training," *Advanced Engineering Informatics,* Nr. 26, p. 439–455, 2012.

[8] L. McAtamney und E. Nigel Corlett, „RULA: a survey method for the investigation of work-related upper limb disorders," *Applied Ergonomics,* Bd. 24, Nr. 2, pp. 91-99, 1993.

[9] A. Musaev, W. Jiangping, L. Zhu, C. Li, Y. Chen, J. Liu, W. Zhan, J. Mei und D. Wang, „Towards in-store multi-person tracking using head detection and track heatmaps," *arXiv:2005.08009v2,* 2 Juli 2020.

[10] F. Courtemanche, P.-M. Léger, A. Dufresne, M. Fredette, É. Labonté-LeMoyne und S. Sénécal, „Physiological heatmaps: a tool for visualizing users' emotional reactions," *Multimed. Tools Appl.,* Nr. 77, p. 11547–11574, 2018.

[11] E. Ferrari, M. Gamberi, F. Pilati und A. Regattieri, „Motion Analysis System for the digitalization and assessment of manual manufacturing and assembly processes," *IFAC PapersOnLine,* Bd. 51, Nr. 11, p. 411–416, 2018.

[12] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei und Y. Sheikh, „OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," *arXiv:1812.08008v2,* 2019.

[13] W. Jastrzębowski, An outline of Ergonomics, or the Science of Work, Warschau: Central Institute of Labour Protection, 1997.

[14] DIN Deutsches Institut für Normung e. V., DIN EN ISO 26800:2011: Ergonomie – Genereller Ansatz, Prinzipien und Konzepte..

[15] J. de Kok, P. Vroonhof, J. Snijders, G. Roullis, M. Clarke, K. Peereboom, P. van Dorst und I. Isusi, Work-related musculoskeletal disorders: prevalence, costs and demographics in the EU, Luxemburg: European Agency for Safety and Health at Work, 2019.

[16] G. David, „Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders," *Occupational Medicine,* Bd. 55, Nr. 3, pp. 190-199, 2005.

[17] O. Karhu, P. Kansi und I. Kuorinka, „Correcting working postures in industry: A practical method for analysis," *Applied Ergonomics,* pp. 199-201, 1977.

[18] Neese Consulting Inc., RULA Employee Assessment Worksheet, USA, 2004.

[19] F. Caputo, A. Greco, E. D'Amato, I. Notaro und S. Spada, „IMU-Based Motion Capture Wearable System for Ergonomic Assessment in Industrial Environment," *Advances in Intelligent Systems and Computing,* pp. 215-225, 2019.

[20] L. Li und X. Xu, „A deep learning-based RULA method for working posture assessment," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* Bd. 63, Nr. 1, pp. 1090-1094, 2019.

[21] S. L. Delp, F. C. Anderson, A. S. Arnold, P. Loan, A. Habib, C. T. John, E. Guendelman und D. G. Thelen, „OpenSim: Open-source software to create and analyze dynamic simulations of movement," *IEEE Transactions on Biomedical Engineering,* Bd. 54, Nr. 11, pp. 1940-1950, 2007.

[22] W. S. McCulloch und W. Pitts, „A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics,* Bd. 5, pp. 115-133, 1943.

[23] L. Wuttke, „Künstliche Neuronale Netzwerke: Definition, Einführung, Arten und Funktion," datasolut GmbH, [Online]. Available: https://datasolut.com/neuronale-netzwerke-einfuehrung/. [Zugriff am 16 März 2021].

[24] M. Träger, A. Eberhart, G. Geldner, A. M. Morin, C. Putzke, H. Wulf und L. H. J. Eberhart, „Künstliche neuronale Netze: Theorie und Anwendungen in der Anästhesie, Intensiv- und Notfallmedizin," *Der Anaesthesist,* Bd. 52, Nr. 11, pp. 1055-1061, 2003.

[25] S. Dörn, „Neuronale Netze," in *Programmieren für Ingenieure und Naturwissenschaftler*, Springer-Verlag, 2018, pp. 89-148.

[26] S. Albelwi und A. Mahmood, „A framework for designing the architectures of deep Convolutional Neural Networks," *Entropy,* Bd. 19, Nr. 242, 2017.

[27] K. O'Shea und R. Nash, „An Introduction to Convolutional Neural Networks," *ArXiv:1511.08458v2,* 2015.

[28] C. Zheng, W. Wu, T. Yang, S. Zhu, C. Chen, R. Liu, J. Shen, N. Kehtarnavaz und M. Shah, „Deep Learning-Based Human Pose Estimation: A Survey," *arXiv:2012.13392v3,* 2020.

[29] J. Wang, X. Long, Y. Gao, E. Ding und S. Wen, „Graph-PCNN: Two Stage Human Pose Estimation with Graph Pose Refinement," *Lecture Notes in Computer Science,* Bd. 12356, pp. 492-508, 2020.

[30] U. Rafi, A. Doering, B. Leibe und J. Gall, „Self-supervised Keypoint Correspondences for Multi-Person Pose Estimation and Tracking in Videos," *Lecture Notes in Computer Science,* pp. 36-52, 2020.

[31] S. Jin, L. Xu, J. Xu, C. Wang, W. Liu, C. Qian, W. Ouyang und P. Luo, „Whole-Body Human Pose Estimation in the Wild," *Computer Vision – ECCV 2020,* pp. 196-214, 2020.

[32] D. Tome, C. Russell und L. Agapito, „Lifting from the Deep: Convolutional 3D Pose Estimation from a Single Image," *Computer Vision and Pattern Recognition,* pp. 2500-2509, 2017.

[33] Y. Cheng, B. Yang, B. Wang, W. Yan und R. T. Tan, „Occlusion-Aware Networks for 3D Human Pose Estimation in Video," *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV),* pp. 723-732, 2019.

[34] M. Fabbri, F. Lanzi, S. Calderara, S. Alletto und R. Cucchiara, „Compressed Volumetric Heatmaps for Multi-Person 3D Pose Estimation," *Computer Vision and Pattern Recognition,* pp. 7204-7213, 2020.

[35] A. Benzine, F. Chabot, B. Luvison, Q. C. Pham und C. Achrd, „PandaNet : Anchor-Based Single-Shot Multi-Person 3D Pose Estimation," *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 6855-6864, 2020.

[36] P. Paudel und K.-H. Choi, „A Deep-Learning Based Worker's Pose Estimation," *Communications in Computer and Information Science,* Bd. 1212, pp. 122-135, 2020.

[37] A. Altieri, S. Ceccacci, A. Talipu und M. Mengoni, „A low cost motion analysis system based on RGB cameras to support ergonomic risk assessment in real workplaces," *ASME 2020 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference,* 2020.

[38] W. Chu, S. Han, X. Luo und Z. Zhu, „Monocular Vision-Based Framework for Biomechanical Analysis or Ergonomic Posture Assessment in Modular Construction," *Journal of Computing in Civil Engineering,* Bd. 34, Nr. 4, 2020.

[39] B. Parsa und A. G. Banerjee, „A Multi-Task Learning Approach for Human Activity Segmentation and Ergonomics Risk Assessment," *arXiv:2008.03014v2,* 2020.

[40] W. Kim, C. Huang, D. Yun, D. Saakes und S. Xiong, „Comparison of Joint Angle Measurements from Three Types of Motion Capture Systems for Ergonomic Postural Assessment," *Advances in Intelligent Systems and Computing,* pp. 3-11, 2020.

[41] Google LLC, „Willkommen bei Colaboratory," [Online]. Available: https://colab.research.google.com/notebooks/intro.ipynb. [Zugriff am 01 Juni 2021].

[42] Google Research, „TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," 2015.

[43] Google Brain, „TensorFlow: A system for large-scale machine learning," *12th USENIX Symposium on Operating Systems Design and Implementation,* pp. 265-283, 2016.

[44] G. Bradski, „The OpenCV Library," *Dr. Dobb's Journal of Software Tools,* 2000.

[45] D. Tome, C. Russell und L. Agapito, „Lifting from the Deep," GitHub, 24 März 2020. [Online]. Available: https://github.com/DenisTome/Lifting-from-the-Deep-release. [Zugriff am 10 Juni 2021].

[46] M. Andriluka, L. Pishchulin, P. Gehler und B. Schiele, „2D Human Pose Estimation: New Benchmark and State of the Art Analysis," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 2014.

[47] C. Ionescu, D. Papava, V. Olaru und C. Sminchisescu, „Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Bd. 36, Nr. 7, 2014.

[48] S. C. Babu, „A 2019 guide to Human Pose Estimation with Deep Learning," Nanonets, 2019. [Online]. Available: https://nanonets.com/blog/human-pose-estimation-2d-guide/. [Zugriff am 2021 Juni 23].

[49] M. Tatariants, „3D Human Pose Estimation Experiments and Analysis," KDnuggets, August 2020. [Online]. Available: https://www.kdnuggets.com/2020/08/3d-human-pose-estimation-experiments-analysis.html. [Zugriff am 23 Juni 2021].

[50] W. Hemmerich, „Winkel zwischen zwei Vektoren," Matheguru, [Online]. Available: https://matheguru.com/lineare-algebra/winkel-zwischen-zwei-vektoren.html. [Zugriff am 14 Juni 2021].

[51] XOVI GmbH, „Was ist eine Heatmap?," [Online]. Available: https://www.xovi.de/was-ist-eine-heatmap/. [Zugriff am 17 Juni 2021].

[52] A. Rosebrock, „Simple object tracking with OpenCV," PyImageSearch, 23 Juli 2018. [Online]. Available: https://www.pyimagesearch.com/2018/07/23/simple-object-tracking-with-opencv/. [Zugriff am 22 Juni 2021].

[53] M.-È. Chiasson, D. Imbeau, K. Aubry und A. Delisle, „Comparing the results of eight methods used to evaluate risk factors associatedwith musculoskeletal disorders," *International Journal of Industrial Ergonomics,* pp. 478-488, 19 August 2012.

[54] Google LLC, „Colab Pro,“ [Online]. Available: https://colab.research.google.com/signup. [Zugriff am 14 Juli 2021].

9 Abbildungsverzeichnis

10 Tabellenverzeichnis

11 Abkürzungsverzeichnis

2D	zweidimensional
3D	dreidimensional
3DSSPP	3D Static Strength Prediction Program
bzw.	beziehungsweise
ca.	Zirka
cm	Zentimeter
CNN	Convolutional Neural Network
COCO	Common Objects in Context
EAWS	European Assembly Work-Sheet
FNN	Feedforward Neural Network
Full HD	Full High Definition
GmbH	Gesellschaft mit beschränkter Haftung
GPR	Graph Pose Refinement
GPU	Graphics Processing Unit, Grafikprozessor
HPE	Human Pose Estimation
ID	Identifikationsnummer
IEA	Internationale Fachgesellschaft für Ergonomie und Arbeitswissenschaft
IMU	Inertial Measurement Unit
LLC	Limited Liability Company (Gesellschaft mit beschränkter Haftung)
LoCO	Learning on Compressed Output
m	Meter
OWAS	Ovako Working Posture Analysing System
PandaNet	Pose estimationAtioN and Dectection Anchor-based Network
PC	Personal Computer
REBA	Rapid Entire Body Assessment
RNN	Recurrent Neural Network
RULA	Rapid Upper Limb Assessment
TPU	Tensor Processing Units, Tensor-Prozessoren
usw.	und so weiter
WLAN	Wireless Local Area Network
z.B.	zum Beispiel