Informatics

Automated Ergonomics Correction with Neural Networks

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Business Informatics

eingereicht von

Omar Drljevic, B.Sc.
Matrikelnummer 12045335

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Univ.-Prof. Dr.-Ing. Dipl.-Ing. Sebastian Schlund
Mitwirkung: Dipl.-Ing. David Kostolani, B.Sc.

Wien, 1. Mai 2024

Omar Drljevic	Sebastian Schlund

TU WIEN Informatics

Automated Ergonomics Correction with Neural Networks

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Business Informatics

by

Omar Drljevic, B.Sc.
Registration Number 12045335

to the Faculty of Informatics

at the TU Wien

Advisor: Univ.-Prof. Dr.-Ing. Dipl.-Ing. Sebastian Schlund
Assistance: Dipl.-Ing. David Kostolani, B.Sc.

Vienna, May 1, 2024

Omar Drljevic   Sebastian Schlund

Erklärung zur Verfassung der Arbeit

Omar Drljevic, B.Sc.

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 1. Mai 2024

Omar Drljevic

Acknowledgements

I extend my heartfelt gratitude to my mentor, David Kostolani, for his invaluable guidance and insights throughout this research journey. His expertise and dedication have been crucial to the success of this project.

I also wish to express my sincere appreciation to Professor Sebastian Schlund for the opportunity to work and learn at the Institute. This experience has been crucial for my professional growth.

My colleagues at the Institute of Management Science deserve special thanks for their constant support and collaboration, which have enriched my research experience and contributed to my professional growth.

I am deeply grateful to my parents for their unwavering support and encouragement, which have been my strength throughout this journey. Their belief in my abilities has been a constant source of motivation.

Lastly, I wish to acknowledge my girlfriend for her understanding, patience, and love, which have been of great help during the most challenging phases of this work.

Kurzfassung

Die steigende Zahl arbeitsbedingter Muskel-Skelett-Erkrankungen bei den Arbeitnehmern in der EU ist ein dringendes Gesundheitsproblem, das die Produktivität der Arbeitnehmer und die wirtschaftliche Stabilität insgesamt erheblich beeinträchtigt. Herkömmliche ergonomische Beurteilungen beinhalten häufig manuelle Bewertungen, die zeitaufwändig sind und häufig zu Verzögerungen bei der Ergreifung von Korrekturmaßnahmen führen. Bei der Entwicklung automatisierter Systeme für die ergonomische Bewertung werden zunehmend Technologien wie Pose Estimation und Motion Capture (Mocap) eingesetzt, um die Ergonomie am Arbeitsplatz in Echtzeit zu bewerten. Bei Technologien zur Posenschätzung werden beispielsweise Kameras und Software eingesetzt, um menschliche Figuren in Bildern und Videos zu erkennen und ihre Körperhaltung anhand bekannter ergonomischer Standards zu bewerten. Diese Methode ermöglicht eine kontinuierliche Bewertung der Ergonomie ohne Unterbrechung des Arbeitsablaufs. Bei Systemen zur Bewegungserfassung hingegen werden Personen mit Sensoren ausgestattet, die die Dynamik ihrer Bewegungen erfassen. Diese Daten werden dann verarbeitet, um ergonomische Risikofaktoren auf der Grundlage der Ausrichtung und Belastung der verschiedenen Körperteile zu bewerten.

Trotz ihrer Schnelligkeit und Effizienz liefern diese automatisierten aber oft Ergebnisse, die von Laien nicht leicht zu interpretieren sind. Sie liefern in der Regel Punktzahlen und statistische Daten, die zwar genau sind, aber keine klaren, umsetzbaren Ratschläge zur Behebung der festgestellten Probleme liefern. Diese Lücke in der Interpretierbarkeit der Daten kann unmittelbare Korrekturmaßnahmen behindern, die für die proaktive Bewältigung ergonomischer Risiken unerlässlich sind. Daher besteht weiterhin ein entscheidender Bedarf an technologischen Weiterentwicklungen, die nicht nur Risiken identifizieren und bewerten, sondern diese Ergebnisse auch in praktische Ratschläge für Anpassungen am Arbeitsplatz umsetzen.

Als Antwort auf diese Herausforderungen stellt diese Studie einen neuen Ansatz. Sie adaptiert Neural Style Transfer Techniken - die ursprünglich für künstlerische Bildtransformationen entwickelt wurden - mit Generative Adversarial Netzwerkkonzepten auf den Bereich der Ergonomie zur Optimierung von Arbeitsplatzhaltungen. Die Methode beinhaltet eine zusammengesetzte Verlustfunktion, die ergonomische, strukturelle und diskriminatorische Verluste integriert, um menschliche Körperhaltungen zu verfeinern. Diese Funktion stellt sicher, dass die veränderten Körperhaltungen nicht nur den ergono-

mischen Sicherheitsstandards entsprechen, sondern auch ein realistisches menschliches Aussehen und strukturelle Integrität bewahren, was das Interesse und Engagement des Publikums weckt.

Dieser neuartige Ansatz ermöglicht die automatische Korrektur von Körperhaltungen und bietet damit einen Vorteil gegenüber herkömmlichen Methoden, die in hohem Maße auf manuelle Eingriffe durch menschliche Ergonomieexperten angewiesen sind. Durch die Nutzung der Möglichkeiten des maschinellen Lernens und neuronaler Netze geht diese Methode das kritische Problem der Korrektur bei bestehenden KI-gestützten Ergonomielösungen an. Sie liefert einen Konzeptnachweis, der ergonomische Risiken bewertet und umsetzbare Anleitungen für Korrekturen liefert, wodurch die Wahrscheinlichkeit von Muskel- und Skeletterkrankungen am Arbeitsplatz verringert wird. Diese Studie bringt den Bereich der Ergonomie voran und legt den Grundstein für künftige Forschungen, die darauf abzielen, ergonomische Bewertungen und Interventionen vollständig zu automatisieren.

Abstract

The rising number of work-related musculoskeletal disorders (WMSDs) in the EU workforce is a pressing health concern that significantly impacts worker productivity and overall economic stability. Traditional ergonomic assessments often involve manual evaluations that are time-consuming and often delayed in delivering corrective measures. The development of automated systems for ergonomic assessment has increasingly harnessed technologies such as pose estimation and motion capture (mocap) to provide real-time evaluations of workplace ergonomics. Pose estimation technologies, for instance, use cameras and software to detect human figures in images and videos and map their posture against known ergonomic standards. This method allows for continuous monitoring of ergonomics without disrupting the workflow. On the other hand, motion capture systems involve fitting individuals with sensors that capture the dynamics of their movement. This data is then processed to evaluate ergonomic risk factors based on the alignment and strain on different body parts.

Despite their speed and efficiency, these automated systems often produce results that are not easily interpretable by non-experts. They tend to offer scores and statistical data that, while accurate, do not provide clear, actionable advice on how to correct identified issues. This gap in the interpretability of the data can hinder immediate corrective actions, which are vital for proactively addressing ergonomic risks. As such, there remains a critical need for technological advancements that not only identify and score risks but also translate these findings into practical advice for workplace adjustments.

In response to these challenges, this study introduces a novel approach. It adapts neural style transfer techniques—originally developed for artistic image transformations—with generative adversarial network concepts to the field of ergonomics for optimizing workplace postures. The method involves a composite loss function that integrates ergonomic, structural, and discriminator losses to refine human postures. This function ensures that modified poses not only meet ergonomic safety standards but also maintain a realistic human appearance and structural integrity, sparking intrigue and engagement in the audience.

This novel approach allows for the automated correction of postures, providing a benefit over traditional methods that rely heavily on manual interventions by human ergonomic experts. By harnessing the capabilities of machine learning and neural networks, this method tackles the critical issue of rectification in existing AI-powered ergonomic solutions.

It provides a proof of concept that assesses ergonomic risks and provides actionable guidance for corrections, thereby reducing the likelihood of MSDs in the workplace. This study advances the field of ergonomics and lays the foundation for future research aimed at fully automating ergonomic assessments and interventions.

Contents

CHAPTER 1

Introduction

1.1 Problem Statement

Chronic work-related injuries, particularly musculoskeletal disorders (MSDs), have emerged as a significant concern in the modern workplace. These disorders not only impact the well-being of employees but also have broader economic implications. MSDs stand as the primary reason for work disability, absenteeism, 'presenteeism' (working while sick), and productivity loss throughout the European Union (EU) member states. The economic ramifications are staggering, with the potential loss in productivity due to MSDs among the working-age population in the EU estimated to be up to 2% of the GDP [Bev15]. This number is concerning, especially if we couple that with consideration of the natural decline in physical capacity and performance with age. Despite various interventions, nearly 150,000 individuals in Austria were granted pensions in 2019 due to diminished working capacity [MC, Str17]. Such statistics underscore the urgency of addressing this issue[1]. [2].

Ergonomics, the scientific discipline dedicated to understanding the interactions between humans and other system elements, offers a promising avenue for intervention. Physical ergonomics focuses explicitly on preventing injuries by thoroughly designing and evaluating workplaces. This sub-discipline emphasizes optimal postures of manual tasks and repetitive movements to minimize injuries, enhance productivity, and reduce errors [3].

Recent advancements in ergonomic research have leveraged machine learning techniques to enhance the analysis of human poses and movements, which are important for mitigating musculoskeletal disorder (MSD) risks. Traditional ergonomic assessment methods such as surveys, expert evaluations, and automated processes are foundational in identifying

[1] https://www.coeh.berkeley.edu/news/difference-between-physical-ergonomics-cognitive-ergonomics

[2] https://auva.at/cdscontent/?contentid=10007.892663&portal=auvaportal

[3] https://www.ergonomics.com.au/what-is-ergonomics/

and correcting risky postures in the workplace. Innovations like Convolutional Neural Networks (CNNs) and Deep Neural Networks (DNNs) are instrumental in evaluating ergonomic metrics, including RULA scores. Additionally, recent developments like the graph-based multi-task learning strategy and the Motion Analysis System (MAS) provide comprehensive assessments, while other techniques generate heatmaps to identify areas of poor ergonomics on factory floors [Dav05, MGHF+21, LLRL21a, KWS22]. Integrating AI technology in ergonomics has shown the potential to enhance precision and reduce the time required for assessments. These automated systems can detect and diagnose ergonomic issues with accuracy, serving as preliminary diagnostic tools to identify areas of concern or potential risk. However, these systems' current inability to provide comprehensive solutions or actionable guidance is a significant limitation. While they can identify problems such as poor posture that may lead to musculoskeletal issues, they do not offer instructions on correcting these issues. This necessitates the intervention of experts for corrective guidance post-diagnosis, indicating that the automation process in ergonomics is still nascent.

The reliance on human expertise remains indispensable, highlighting that the automation of the ergonomic assessment process has yet to be achieved. The real value for workers and stakeholders would be realized from a system that not only identifies the issues but also provides tangible solutions to rectify them. Until such advancements are made, the role of AI in ergonomics will likely remain supplementary to traditional methods like RULA, REBA, NIHOS or EAWS rather than a complete replacement. This underscores the need for continued development in AI-driven ergonomic solutions that can offer both diagnosis and corrective guidance.

In addressing this evident gap in the automation of ergonomic assessments, this thesis is inspired by the innovative application of the neural style transfer technique. Neural style transfer is a technique in deep learning that blends the content of one image with the style of another image. The process optimizes a loss function to minimize differences in content and style between the synthesized and source images [IZZE17, KKS+23, SJJ+21]. The concept of neural style transfer, originally developed for image processing, can also be applied in ergonomics. This thesis demonstrates how neural style transfer can effectively superimpose correct pose sequences over original sequences featuring unergonomic poses. The methodology involved defining a custom loss function that quantifies the alignment between generated data and a target pose sequence that exemplifies desirable ergonomic attributes. This loss is then minimized using gradient-based optimization methods such as gradient descent.

The proposed approach captures and refines a worker's pose sequence towards an ideal ergonomic posture. The final output identifies discrepancies between current and optimal postures, providing visual guidance for necessary corrections. This approach seeks to bridge the current gap by offering a comprehensive solution: identifying the ergonomic issue, evaluating its implications, and providing actionable guidance for correction, all through the power of advanced neural networks. This could give a new perspective on how we approach and address workplace ergonomics, making the process genuinely automated and highly effective.

1.2 Research Aim and Questions

This thesis seeks to develop a novel implementation of postural optimization in automated ergonomic assessment through the application of neural style transfer techniques. Specifically, the research aims to adapt and harness the capabilities of neural style transfer, a technique traditionally used in image processing, to the context of ergonomic assessments, thus optimizing workplace postures for better health and productivity outcomes. By integrating this approach, the project addresses a gap in current ergonomic practices, which often rely heavily on manual assessments and are limited in their ability to provide corrective, adaptive solutions. The research questions focus on evaluating the effectiveness of this novel application:

RQ1: What constitutes optimal ergonomics, and which metrics describe ergonomics?

RQ2: To what extent can neural network optimization be employed to optimize posture while preserving the intended action?

RQ3: Which hyperparameter combination constitutes ergonomic optimization qualitatively?

1.3 Methods

1.3.1 Literature and State of the Art Review

The literature review of this thesis examined ergonomics, starting with fundamental ergonomic principles and progressing to scoring techniques such as RULA and REBA, which assess the risk of musculoskeletal disorders in workplace settings. Additionally, the review studies the conditions and challenges that precipitate ergonomic issues in various industries. To support the approach of this research, the review extends into computational methodologies, including neural style transfer, neural optimizations, and generative adversarial networks (GANs).

1.3.2 Prototype Development

Developing the prototype for this thesis was driven by an extensive literature review across various fields, including ergonomics, optimization, and neural networks. The absence of existing models integrating neural style transfer for ergonomic assessments prompted a unique approach. This comprehensive examination of existing methodologies, especially those utilized in image processing, informed the strategies employed in this project. Each stage of the prototype's development involved rigorous iterative testing and refinement, an approach to combine neural network technologies with ergonomic assessments.

1.3.3 Prototype Evaluation

The prototype developed in this thesis was evaluated through objective measurements and subjective assessments. Subjectively, the effectiveness of the newly generated postures was analyzed by visually inspecting the improvements in ergonomic alignment as suggested by the neural style transfer outputs. Objectively, the prototype's success was quantified by measuring the percentage reduction in ergonomic risk scores, such as RULA or REBA scores, before and after applying the proposed AI enhancements. This dual approach allowed for an assessment of the prototype, providing insights into the postures' perceptual quality and the tangible ergonomic benefits achieved through the neural-based optimization process.

1.4 Structure of the Work

Seven sections make up this thesis, which as a whole addresses the study objectives and seeks to provide an extensive understanding of the topic. The main points discussed in each section are outlined in the summary that follows:

Introduction: Sets the stage by addressing the issue of musculoskeletal disorders and the significance of ergonomic assessments.

Theoretical Foundations: Explores fundamental ergonomic concepts and the physiological and biomechanical bases of ergonomics, reviews traditional and emerging assessment methodologies, and establishes a foundational knowledge base to support advanced technological exploration.

State of the Art: Reviews advancements in ergonomic techniques and artificial intelligence, focusing on their applicability and limitations in the context of ergonomic assessments and discuss current machine learning models, especially CNNs and DNNs, in detail.

Implementation: Details the technical development of the prototype, including challenges faced and solutions found but also describes the integration of neural style transfer into ergonomic posture evaluation and the technical framework developed.

Results and Evaluation: Presents both subjective and objective evaluations of the prototype and also details the effectiveness of neural style transfer in enhancing ergonomic postures and quantitatively reducing ergonomic risk scores.

Discussion: Interprets the results, assessing the practical implications and feasibility of AI-enhanced ergonomic assessments. This chapter explores the potential for future enhancements and broader applications within ergonomics.

Conclusion: Summarizes the key contributions and reaffirms the innovative integration of AI in ergonomics while discussing the potential impact on improving workplace health and productivity and suggests directions for future research.

CHAPTER 2

Theoretical Foundations

2.1 Understanding Musculoskeletal Disorders

Work-related Musculoskeletal Disorders (WMSDs) stand as a significant public health issue, especially pronounced in industrialized nations, including the European Union and the United States, where the frequency of WMDs is far greater than circulatory or chronic respiratory conditions [MBI18]. WMSDs encompass a range of injuries and afflictions impacting the muscles, nerves, tendons, joints, and other parts of the human musculoskeletal system [1]. The nature of WMSDs varies a significant amount, but a substantial body of evidence points to occupational hazards as the primary cause. Specifically, WMSDs arise from or are exacerbated by the workplace environment and the nature of the work performed. This includes, but is not limited to, repetitive, forceful tasks, awkward postures, and sustained exposure to vibrations — all of which are associated with the development of conditions such as sprains, strains, carpal tunnel syndrome, and chronic back pain .

The statistical sources paint a stark picture: more than half of the adult population in the United States suffers from WMSDs, with low back and chronic joint pains being among the most reported complaints. Such disorders have shown an increase in prevalence with age, particularly affecting individuals over the age of 65. Despite advancements in workplace ergonomics and a strategic focus on prevention within the European Union, data has unfortunately revealed a persistent, unmitigated exposure to WMSD risk factors across most sectors. This is compounded by the demographic shift towards an older workforce, which is inherently at a higher risk for developing WMSDs[JCDG]. However, it can be challenging to interpret the available data on WMSD prevalence and its causes. This difficulty is partly due to potential biases in reporting; women are generally more

[1] https://www.cdc.gov/workplacehealthpromotion/health-strategies/musculoskeletal-disorders/index.html

likely to report MSDs, whereas men may underreport these conditions, skewing prevalence data and complicating the formulation of accurate conclusions [MBI18, JCDG]. The persistent nature of WMSDs, coupled with the lack of a significant decrease in risk exposure, highlights the ongoing challenges in effectively countering the root causes of these disorders.

To address this significant health concern, there has been a growing emphasis on developing and implementing comprehensive workplace health programs tailored to mitigate the risks associated with WMSDs. These initiatives often integrate ergonomic assessments, employee training, and modifications of work practices to reduce the occurrence and severity of musculoskeletal disorders [MBI18, JCDG]. Such programs also involve proactive health promotion strategies, including regular exercise and stress management, which have shown effectiveness in enhancing overall employee well-being and reducing the incidence of WMSDs. Moreover, technological advancements, such as automation and assistive devices, are increasingly being employed to alleviate the physical demands on workers. Additionally, there is a concerted effort to improve the reporting and documentation of WMSDs, aiming for a more accurate and inclusive understanding of the issue. The economic and social impacts of Musculoskeletal Disorders (MSDs) underscore the pressing need for the comprehensive workplace health programs described earlier. These initiatives aim to improve individual well-being and reduce the incidence of MSDs. They also serve as crucial measures in combating the significant economic burden and work absences caused by these disorders in the European Union.

Musculoskeletal disorders (MSDs) present a substantial economic challenge to the European Union (EU), as they are a significant cause of work absence and long-term incapacity. The ageing workforce, further burdened by the rising prevalence of chronic diseases, intensifies this challenge. With an older population remaining in employment longer due to the pension crisis, the incidence of MSDs is set to increase, inevitably impacting economic productivity. The statistics are stark: MSDs account for half of all work absences and 60% of permanent incapacity in Europe, leading to an estimated 1-2% reduction in the gross domestic product (GDP) of individual member states [Bev15]. As the workforce struggles with the physical limitations brought on by these disorders, there is a significant knock-on effect on the economy, manifesting in reduced labour output and increased healthcare costs, as well as compensation claims that can amount to 40% of worker compensation costs in some EU countries[LGBD+14]. Furthermore, despite various interventions, nearly 150,000 individuals in Austria were granted pensions in 2019 due to diminished working capacity [2] [3] [Str17].

The social implications of MSDs are equally profound. Premature exit from the labour market due to such health issues can precipitate a cascade of negative consequences, ranging from individual financial loss and increased risk of poverty to broader social exclusion. This interplay further complicates early return to work and prolongs the

[2] `https://www.wifo.ac.at/publikationen/publikationssuche?detail-view=yes&publikation_id=68042`

[3] `https://auva.at/cdscontent/?contentid=10007.892663&portal=auvaportal`

duration of work-related disability. Moreover, the impact extends beyond the individual to their families and carers, often resulting in disrupted work lives and diminished productivity for those providing care. Effectively managing this issue is vital for achieving the European Commission's and national governments' objectives as they work to balance these connected health and economic concerns.

2.2 Basics of Ergonomics

The scientific field of ergonomics studies how people interact with other system components. It is a profession that uses theory, concepts, data, and methodologies to create systems that maximise human well-being and overall system performance [4]. A critical aspect of ergonomics, particularly relevant in workplace health and productivity discussions, is its role in mitigating the risks associated with Musculoskeletal Disorders (MSDs). The essence of ergonomics, with its commitment to tailoring workspaces and tools to fit the user, inherently includes strategies to prevent MSDs, which are conditions that can affect the muscles, nerves, blood vessels, ligaments, and tendons. By considering factors such as body size, strength, sensory abilities, psychological attitudes, and the potential ergonomic risk factors that can lead to MSDs, ergonomics strives to create spaces that are not only safe and efficient but also conducive to overall well-being, thereby reducing the incidence and impact of these disorders [DMM05].

The methodology of ergonomics is rooted in a synergy of several scientific fields. It employs anthropometry to understand the diverse physical dimensions of the human population, biomechanics to grasp the mechanical aspects of muscular movements, and environmental physics to assess the impact of external conditions like noise and temperature on the human body. Furthermore, ergonomics draws upon applied psychology to address individual differences in skill and behaviour and social psychology to consider group dynamics and communication within work environments. This multidisciplinary approach is pivotal in creating workstations that prioritize the employee's comfort and health, aiming to eradicate the risk of injury by adapting the work to the worker's needs. This proactive and human-centric design philosophy is what sets ergonomics apart as a crucial element in optimizing both individual health and the overall system performance [5].

Ergonomic assessments, essential for ensuring safety and efficiency in various work environments, incorporate a range of tools and techniques. A detailed survey of Certified Professional Ergonomists (CPEs) provides a comprehensive view of the practices in ergonomic assessments [DMM05]. The survey's findings highlight the frequent use of traditional tools such as tape measures, video cameras, stopwatches, and digital cameras, which are instrumental in documenting physical aspects of work environments for ergonomic analysis.

Observational methods, focusing on manual materials handling, are commonly employed alongside these tools. Additionally, direct measurement tools like pinch and grip dy-

[4] https://ehs.unc.edu/topics/ergonomics/
[5] https://www.ergonomics.com.au/what-is-ergonomics/

namometers and push/pull gauges are popular for assessing physical exertions in work settings. The survey also emphasizes the significant role of digital tools, including software and checklists, in ergonomic assessments. These tools, often augmented with anthropometric data, offer structured approaches for evaluating work environments [DMM05].

Advancements in ergonomic assessment devices have led to the integration of more sophisticated tools such as Inertial Measurement Units (IMUs), motion sensing imaging systems, Galvanic Skin Response(GSR), and Eye tracking devices [TLS+17a]. IMUs combine accelerometers, gyroscopes, and occasionally magnetometers to measure both static and dynamic aspects of movement, providing detailed data on physical postures and motions. Their application in real-time ergonomic assessments, particularly in industrial settings, demonstrates their versatility and effectiveness. However, challenges like magnetic disturbances from the units have been observed [TLS+17a].

Motion devices, initially developed for gaming, have been adapted for ergonomic assessments due to their 3D mapping and motion recognition capabilities. It has been utilized in various studies to analyze work-related postures and movements, effectively detecting ergonomic risk factors in diverse settings. Despite its innovative approach, these devices are not without limitations, including sensitivity to light conditions and tracking accuracy of joint movements.

Galvanic Skin Response (GSR) Sensors, also known as skin conductance, measure the electrical conductance of the skin, which varies with its moisture level. In ergonomic assessments, GSR sensors can be particularly useful in measuring the physiological responses of individuals to their work environment. This includes stress responses to physical strain or awkward postures. By monitoring these responses, ergonomists can gain insights into the physical demands of tasks and the potential stress they place on workers, which might not be evident through visual observation alone. This data can be crucial in designing workplaces that reduce physical stress and enhance worker comfort and safety [OLC+23].

Eye-tracking technology is another tool being used in ergonomic studies. This technology tracks where and how long a person looks at various elements within their environment. In the context of physical ergonomics, eye-tracking can provide valuable information about a worker's interaction with their environment. For instance, it can identify issues with workstation design that may cause unnecessary head, neck, or eye movements, leading to strain or fatigue. Eye-tracking can also help in understanding how workers process visual information in their environment, which is crucial for optimizing the layout of controls, displays, and other critical elements in a workspace.

The integration of traditional tools with advanced devices such as Inertial Measurement Units (IMUs), Galvanic Skin Response (GSR) sensors, and motion cameras marks a significant advancement in ergonomic assessment practices. This approach addresses the limitations inherent in each individual tool, providing a more robust and precise evaluation of ergonomic risks. This is particularly valuable in complex work environments where the dynamics of physical strain and stress are more intricate [TLS+17a].

Expanding on this, the fundamental rationale behind employing such a mix of traditional and modern tools is the need for quantification. The primary objective is to achieve a more precise and accurate quantification of ergonomic risks. This enhanced quantification allows for better understanding and mitigation of these risks, leading to improved workplace safety and employee well-being. Essentially, the use of a combination of sensors and assessment methods is not merely for observation but for quantifying the risks in measurable terms, which is vital for effective ergonomic management.

In conclusion, the ergonomic assessment field has witnessed a significant transformation, blending traditional methods with innovative technologies. Insights from the survey of CPEs [DMM05] and recent developments in ergonomic devices [TLS+17a] underscore this evolving landscape, pointing towards a future where comprehensive and precise assessments are more attainable, enhancing safety and efficiency in work environments

2.3 Quantifying Postural Risks

In the workplace health and safety domain, a key focus is the understanding and management of ergonomic risks, with a particular emphasis on musculoskeletal disorders (MSDs). This connection between MSDs, ergonomics, and various assessment techniques is critical in ensuring a healthy working environment.

The rationale behind emphasizing ergonomic risks is clear: if left unchecked, poor body posture and forceful working methods can result in irreversible damage to body tissues. To address these concerns, it is essential to classify and analyze various postures and forces. This classification helps identify specific instances or locations where ergonomic issues are most prevalent, thus paving the way for timely and effective interventions. These postural issues can often be traced back to poorly designed jobs or workstations.

In practice, ergonomic assessments can encompass a range of factors, including postural analysis, forces exerted during work, environmental conditions like light and noise levels, and even the design of tools and equipment. For example, an assessment might analyze the posture of a worker at a computer workstation, examining factors such as chair height, screen position, and keyboard placements [Par00, BW20]. Similarly, in a manufacturing setting, the focus might be on the forces exerted during lifting or the repetitive movements required by a specific task.

Various assessment techniques have been devised to gain insights into these ergonomic challenges. Some methods, like the Body Part Discomfort scale or questionnaires, adopt a subjective approach, relying on individual perceptions and feedback. These self-assessment techniques allow workers to report their own experiences of discomfort or strain, which can be particularly useful for identifying less obvious ergonomic issues [KBH17].

Most postural assessment techniques can vary in their application depending on the measurement method. Objective methods like RULA (Rapid Upper Limb Assessment), REBA (Rapid Entire Body Assessment), OWAS (Ovako Working Posture Analysing System), the NIOSH Lifting Equation, and EWAS (The European Assembly Worksheet) offer standardized frameworks for ergonomic evaluation [TLS+17b]. These expert-based techniques, such as the RULA, provide systematic and quantifiable means to assess body posture, assigning scores to indicate ergonomic risk. They can be applied in event- and time-driven contexts, adapting to immediate or long-term postural analysis, thus leading to targeted and effective ergonomic interventions. Consequently, event-driven postural assessment techniques exhibit heightened sensitivity to variations, in contrast to their time-driven counterparts. Meanwhile, the time-driven postural assessment methods are more abstract and generalized [BHSS07].

The need for quantification in ergonomic assessment stems from the desire to understand and mitigate risks objectively. Quantifying ergonomic risks allows organizations to prioritize interventions, allocate resources effectively, and track the impact of ergonomic improvements over time. It transforms the somewhat subjective nature of ergonomic complaints into measurable data that can guide informed decisions.

These techniques are theoretical concepts and practical tools that can foster healthier work environments, mitigate risks, and improve worker well-being. The ultimate aim is to establish a work environment that promotes health and productivity, reduces the occurrence of MSDs, and enhances the quality of work life for all employees. The following sections delve into these techniques.

RULA (Rapid Upper Limb Assessment)

RULA is an event-driven postural assessment method meticulously designed to identify ergonomic risks in various work tasks. Developed by McAtamney and Corlett in early 1993 [MC93], RULA evaluates tasks' posture, force, and movement dynamics, emphasising the upper body. Its scoring system, which ranges from 1 to 7, provides insight into the ergonomic risk posed by specific tasks. These scores are conveniently grouped into four action levels or categories, each indicating a timeframe within which risk control measures should ideally be initiated. Such tasks often encompass screen-based or computer tasks, manufacturing, or retail tasks where workers sit or stand in static positions.

The method for calculating the RULA score, detailed in 2.1, operates as an event-driven postural assessment methodology, where evaluations are conducted in response to specific actions or changes in posture during work tasks. This targeted approach allows for accurate assessments of ergonomic risks at critical moments. This implies that during the computation of the RULA score, an ergonomic specialist or analytical software evaluates individual postures at discrete moments.

RULA applications are multifaceted and play an important role in various industries. Its primary uses include measuring musculoskeletal risk, which is often a part of a comprehensive ergonomic investigation. This tool also compares musculoskeletal strain between current and modified workstation designs, helping optimise workspaces. Additionally, RULA is instrumental in evaluating outcomes that may include metrics like

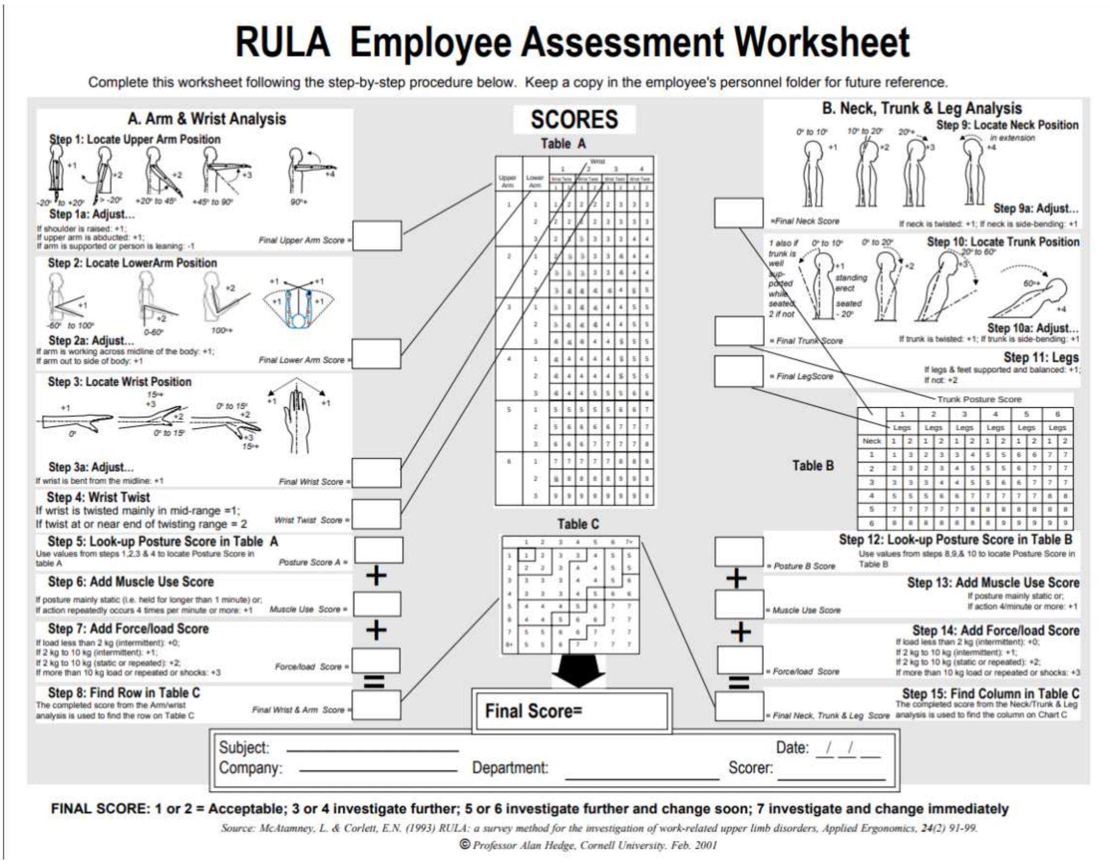


Figure 2.1: RULA Sheet

productivity or the compatibility of specific equipment with the workforce. Another significant application of RULA is educating workers about the musculoskeletal risks associated with different working postures. Due to its straightforward methodology and broad scope of applications, RULA has become an indispensable tool, fostering proactive ergonomic interventions and enhancing workplace safety and efficiency [MC93].

REBA (Rapid Entire Body Assessment)

Hignett and McAtamney introduced REBA in the UK in 2000, recognizing a gap in postural analysis tools. Specifically, they aimed to address the nuanced and variable working positions prevalent in health care, such as handling animate loads, and in other service industries [HM00]. Unlike RULA, which concentrates primarily on the upper body, REBA provides a holistic assessment, encompassing upper and lower body dynamics. However, as with RULA, REBA is an event-driven postural assessment method. This assessment tool collates data on various parameters, encompassing body posture, exerted forces, nature of movements, repetition rate, and the type of coupling. An aggregate REBA score is then computed, indicating the associated ergonomic risk and the immediate

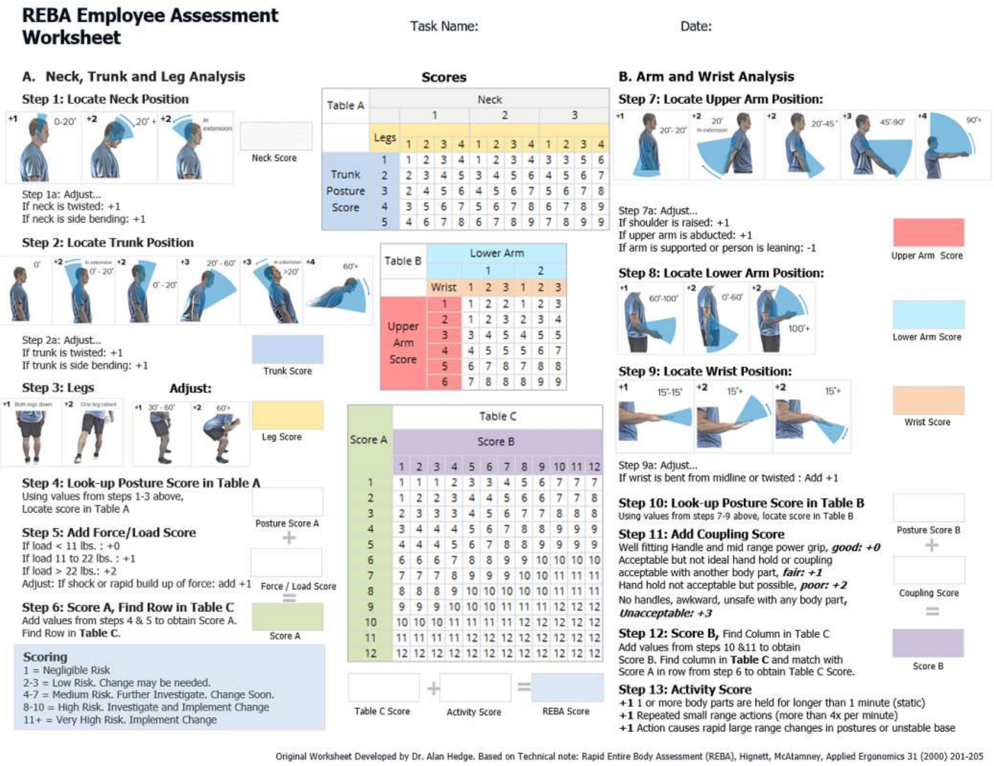
Table 2.1: Steps of RULA Assessment

Step	Analysis Step Description	Area of Analysis
1	Locate Upper Arm- Angel	Arm and Wrist Analysis
2	Locate Lower Arm Angel	
3	Locate Wrist Angel	
4	Locate Wrist Twist Angel	
5	Determine Posture Score A	
6	Add Muscle Use Score	
7	Add Force/Load Score	
8	Find Row in Posture Score C	
9	Locate Neck Angel	Neck, Trunk and Leg Analysis
10	Locate Trunk Angel	
11	Determine Legs Condition	
12	Determine Posture Score B	
13	Add Muscle Use Score	
14	Add Force/Load Score	
15	Find Column in Posture Score C	
16	Determine Final Score (1-7)	

need for remedial action. The foundational principles of REBA are deeply rooted in established ergonomic frameworks, drawing from the tenets of RULA, OWAS, and NIOSH. A cornerstone of this assessment method is the recognition of the functional anatomically neutral posture, as defined by the American Academy of Orthopedic Surgeons in 1965. As postures deviate from this neutral stance, the risk score proportionately escalates. REBA provides comprehensive tables, facilitating the conversion of distinct posture combinations into a singular score, encapsulating the musculoskeletal risk level. These scores are subsequently categorized into five distinct action levels, each offering guidance on the immediacy of interventions to mitigate or eliminate the assessed posture's risk. REBA's applicability is vast, making it an invaluable tool in ergonomic assessments, especially when evaluating full-body involvement, varied postures, or the handling of animate or inanimate loads, regardless of their frequency. The method for calculating the RULA (Rapid Upper Limb Assessment) score is detailed in Figure 2.2.

OWAS (Ovako Working Posture Analysing System)

The OWAS (Ovako Working Posture Analysis System) method, developed by O Karhu, P Kansi, and I. Kuorinka in 1977, is an ergonomic assessment tool for analysing and categorising work postures. Unlike REBA (Rapid Entire Body Assessment) and RULA (Rapid Upper Limb Assessment), which are event-driven, OWAS adopts a time-driven approach. This means that ergonomic experts or specialised analytical software continuously observe a sequence of postures over time [KKK77].



REBA Employee Assessment Worksheet

Task Name:

Date:

A. Neck, Trunk and Leg Analysis

Step 1: Locate Neck Position

+1 0-20° +2 20° + +2 in extension

Neck Score

Step 1a: Adjust...
If neck is twisted: +1
If neck is side bending: +1

Step 2: Locate Trunk Position

+1 0° +2 in extension 0°-20° +2 0°-20° +3 20°-60° +3 in extension >20° +4 60°+

Step 2a: Adjust...
If trunk is twisted: +1
If trunk is side bending: +1

Trunk Score

Step 3: Legs **Adjust:**

+1 Both legs down +2 One leg raised +1 30°-60° +2 60°+

Leg Score

Step 4: Look-up Posture Score in Table A
Using values from steps 1-3 above,
Locate score in Table A

Posture Score A

Step 5: Add Force/Load Score
If load < 11 lbs. : +0
If load 11 to 22 lbs. : +1
If load > 22 lbs.: +2
Adjust: If shock or rapid build up of force: add +1

Force / Load Score

Step 6: Score A, Find Row in Table C
Add values from steps 4 & 5 to obtain Score A.
Find Row in **Table C.**

Score A

Scoring
1 = Negligible Risk
2-3 = Low Risk. Change may be needed.
4-7 = Medium Risk. Further Investigate. Change Soon.
8-10 = High Risk. Investigate and Implement Change
11+ = Very High Risk. Implement Change

Scores

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

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

Table C: Score A \ Score B	1	2	3	4	5	6	7	8	9	10	11	12
1	1	1	1	2	3	3	4	5	6	7	7	7
2	1	2	2	3	4	4	5	6	6	7	7	8
3	2	3	3	3	4	5	6	7	7	8	8	8
4	3	4	4	4	5	6	7	8	8	9	9	9
5	4	4	4	5	6	7	8	8	9	9	9	9
6	6	6	6	7	8	8	9	9	10	10	10	10
7	7	7	7	8	9	9	9	10	10	11	11	11
8	8	8	8	9	10	10	10	10	10	11	11	11
9	9	9	9	10	10	10	11	11	11	12	12	12
10	10	10	10	11	11	11	11	12	12	12	12	12
11	11	11	11	11	12	12	12	12	12	12	12	12
12	12	12	12	12	12	12	12	12	12	12	12	12

Table C Score + Activity Score = REBA Score

B. Arm and Wrist Analysis

Step 7: Locate Upper Arm Position:

+1 20°-20° +2 20° in extension +2 20°-45° +3 45°-90° +4 90°+

Step 7a: Adjust...
If shoulder is raised: +1
If upper arm is abducted: +1
If arm is supported or person is leaning: -1

Upper Arm Score

Step 8: Locate Lower Arm Position:

+1 60°-100° +2 0°-60° +2 100°+

Lower Arm Score

Step 9: Locate Wrist Position:

+1 15°-15° +2 15°+ +2 15°+

Wrist Score

Step 9a: Adjust...
If wrist is bent from midline or twisted : Add +1

Step 10: Look-up Posture Score in Table B
Using values from steps 7-9 above, locate score in Table B

Posture Score B

Step 11: Add Coupling Score
Well fitting Handle and mid range power grip, ***good: +0***
Acceptable but not ideal hand hold or coupling acceptable with another body part, ***fair: +1***
Hand hold not acceptable but possible, ***poor: +2***
No handles, awkward, unsafe with any body part, ***Unacceptable: +3***

Coupling Score

Step 12: Score B, Find Column in Table C
Add values from steps 10 &11 to obtain Score B. Find column in **Table C** and match with Score A in row from step 6 to obtain Table C Score.

Score B

Step 13: Activity Score
+1 1 or more body parts are held for longer than 1 minute (static)
+1 Repeated small range actions (more than 4x per minute)
+1 Action causes rapid large range changes in postures or unstable base

Original Worksheet Developed by Dr. Alan Hedge. Based on Technical note: Rapid Entire Body Assessment (REBA), Hignett, McAtamney, Applied Ergonomics 31 (2000) 201-205

Figure 2.2: REBA Sheet [HM00]

In OWAS, postures are meticulously distinguished and categorised based on four aspects for the back, three for the arms, and seven for the legs. Additionally, the weight of the load handled is classified into three distinct categories. This categorisation allows for a systematic evaluation of postures at consistent intervals, providing a detailed analysis of the overall body posture. This is achieved by amalgamating data from these individual body parts.

A unique feature of OWAS is its use of a four-digit code to represent postures. This code encapsulates data about the back, upper limbs, lower limbs, and the required force. The posture combinations identified are then stratified into four action categories. These categories are based on expert assessments of the potential health ramifications associated with each specific posture or combination.

The OWAS method stands out for its detailed and systematic approach to posture analysis. Focusing on continuous observation of postures and categorising them comprehensively provides a nuanced understanding of the ergonomic risks in various work environments, as seen in Figure 2.6.

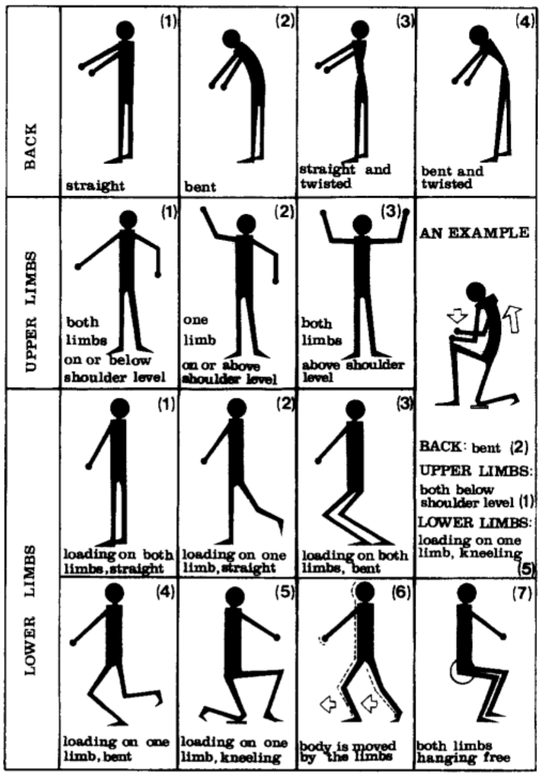


Figure 2.3: List of items classified by OWAS (the Ovako Working Posture Analysing System). A code number is given for each item. Each posture can be described with a three-digit code. The example in the right side of the figure, thus, can be represented by the code 215 [KKK77]

NIOSH Lifting Equation Explanation

The revised NIOSH Lifting Equation, introduced in 1994, stands as a critical instrument in the field of ergonomics, specifically designed to evaluate the safety parameters of manual material handling tasks [WPAG94]. It systematically quantifies the individual's lifting capacity by examining a spectrum of eight fundamental factors: the weight of the load, its horizontal distance from the body, the vertical height of the lift from the floor, the total vertical distance traversed during the lift, the degree of twisting imposed on the back, the frequency and duration of the lifting task, and the quality of the object's grip.

The Recommended Weight Limit (RWL) concept is at the core of this equation. The RWL determines the maximum safe weight a worker can lift under optimal conditions. It is not just a static number; it varies based on the specific conditions of each lifting task, as influenced by the abovementioned factors. The RWL is calculated through a formula that integrates these factors, providing a tailored assessment of each lifting scenario. This personalized approach ensures that the safety guidelines are relevant and practical for a wide range of real-world lifting tasks, significantly reducing the risk of injuries related to manual material handling. The RWL is derived using the following formula:

$$RWL = LC \times HM \times VM \times DM \times AM \times FM \times CM \tag{2.1}$$

Each component of this equation represents a variable that can influence lifting safety:

- RWL is the Recommended Weight Limit.
- LC stands for the Load Constant, which is 23.13 kg (51 pounds) for the 40th percentile female.
- HM is the Horizontal Multiplier, reflecting the horizontal reach required to grasp the load.
- VM is the Vertical Multiplier, indicating the vertical distance from the floor to the load's starting point.
- DM is the Distance Multiplier, accounting for the range of motion during the lift.
- AM is the Asymmetry Multiplier, quantifying the rotational demand on the body.
- FM is the Frequency Multiplier, denoting the rate of lifting activity.
- CM is the Coupling Multiplier, assessing the ease of holding the load.

The Load Constant (LC) is particularly significant as it establishes a baseline for ideal lifting conditions based on the strength of the 40th percentile of females, allowing a safe lift of 23.13 kg. When lifting conditions deviate from the ideal—due to factors like excessive load distance from the body or an awkward starting height—the equation's multipliers come into play, adjusting the RWL to prevent injuries by compensating for the additional strain on the lifter's body.

Precision in measurement is crucial for accurately applying the NIOSH Lifting Equation. The horizontal distance must be measured from the midpoint between the ankles to the object, while the vertical distance is gauged from the floor to the lifter's knuckles when in the lifting position. These measurements are essential for computing the asymmetry angle and the travel distance of the lift, both of which are critical for establishing safe lifting practices.

Moreover, the equation considers the lifting frequency, averaged over 15 minutes and categorized the duration of lifting into short, moderate, and long intervals. These temporal factors are vital in assessing the cumulative load on the body, thereby aiding in determining necessary rest periods to prevent worker fatigue and overexertion [WPAG94].

To contextualize the RWL, it is compared against the actual Load Weight (LW) through the Lifting Index (LI), which is calculated as follows:

$$LI = \frac{\text{Load Weight}}{\text{Recommended Weight Limit}} = \frac{L}{RWL} \tag{2.2}$$

The LI indicates the physical stress level associated with a lifting task. A LI value close to or less than 1 suggests that the lift is within ergonomic safety limits. Conversely, a LI significantly greater than 1 indicates a lift that may pose ergonomic risks, signalling the need for task redesign or intervention.

Lifting Index Value	**Safety level**
LI < 1	Safe and Healthy for all workers
1 < LI < 2	Safe but can begin to be physically stressful
3 < LI	Significant levels of physical stress and health concerns

Table 2.2: Safety Levels Based on Lifting Index Value

The NIOSH Lifting Equation presents a structured approach for assessing and amending lifting tasks, which is crucial in identifying risk factors linked with these activities. It is important to note that the NIOSH Lifting Equation is specifically tailored for lifting tasks, with its applicability limited to this area. Its precision and reliability make it a valuable asset in occupational health. However, its use is confined to workplaces that involve lifting, highlighting the need for other specialized tools and techniques for different types of workplace hazards.

EWAS (The European Assembly Worksheet)

The European Assembly Worksheet (EAWS) stands as a notable document created by a consortium of ergonomics specialists from the Institute of Ergonomics at Darmstadt University of Technology, alongside the International MTM Directorate and the MTM-Institut [SCBB13]. Its primary purpose is to serve as a preventative measure against musculoskeletal disorders (MSDs), which are prevalent in industrial assembly roles and can lead to significant work-related disabilities. The EAWS provides a robust framework with detailed guidelines across various critical sections, including but not limited to optimal working postures, the exertion of action forces, the handling of materials manually, and the repetitive movements of the upper limbs, all designed to improve the ergonomics of assembly tasks and reduce the incidence of MSDs.

The EAWS operates on a colour-coded three-zone rating system akin to traffic lights: green for low risk, yellow for moderate risk, and red for high risk. This evaluation spans four key sections: working postures, action forces (both whole body and hand-finger system), manual material handling, and the impact of repetitive movements on upper limbs ass seen in Figures 2.4 and 2.5. Within each section, points are allocated based on several criteria. These include the extent of deviation from ergonomically favourable postures, the level of force required for tasks, the weight of objects handled, and the duration and frequency of repetitive tasks. The points accumulated in each section fall within the color-coded system: 0 to 33 points for green, 34 to 66 points for yellow, and 67

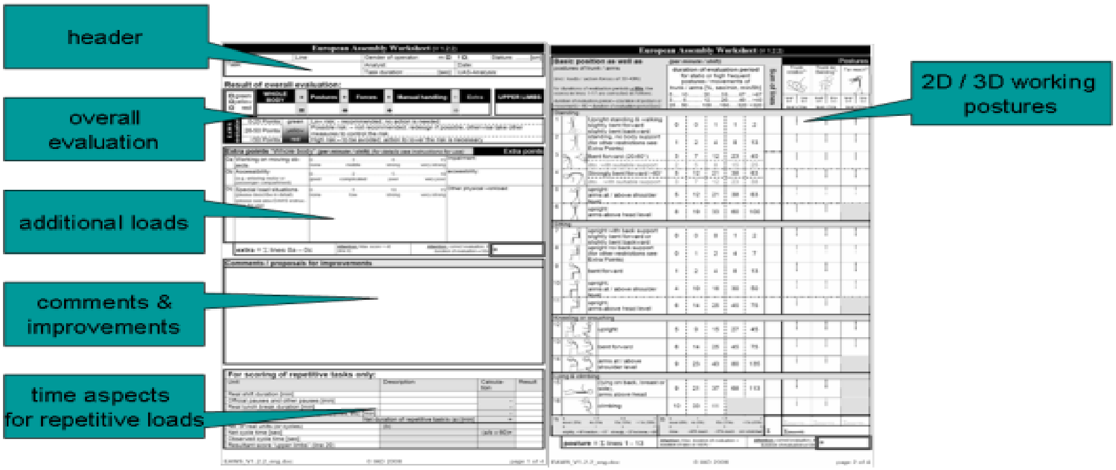


Figure 2.4: EWAS structure and sections [SCBB13]

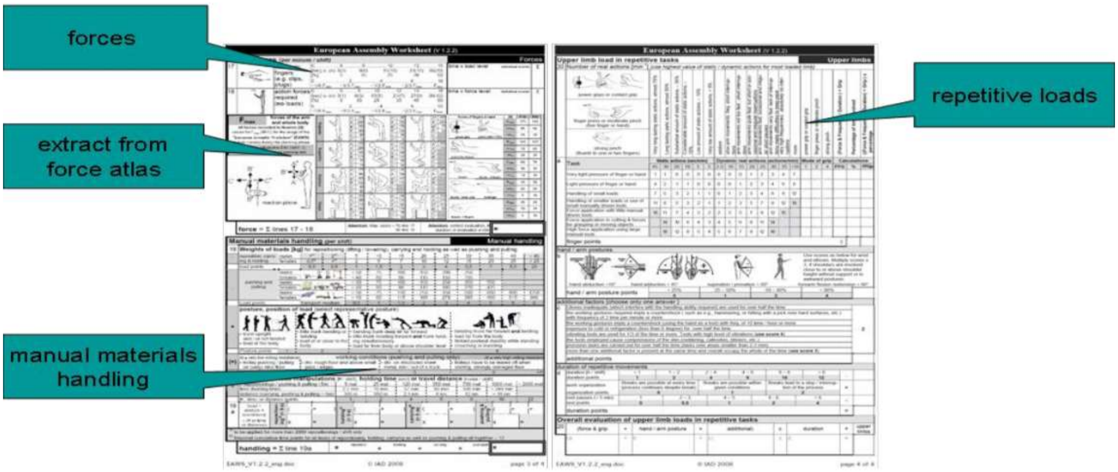


Figure 2.5: EWAS structure and sections II [SCBB13]

to 100 points for red. The thresholds between these zones are transitional, indicating a similar level of risk for scores near the boundary [SCBB13]. The overall ergonomic risk of a task is determined by the highest score in any of the sections, ensuring comprehensive risk assessment. For instance, if only one aspect of a task scores in the red zone, the entire task is classified as high risk. This methodical scoring system enables employers to identify and address high-risk tasks, reducing the likelihood of musculoskeletal disorders (MSDs) and enhancing workplace safety and efficiency.

2.4 Neural Networks and Image Processing

Incorporating neural networks into ergonomics research, particularly in video data analysis, marks a significant leap forward. Tools like pose estimation, powered by these computational models, provide deeper insights into human posture and movement. This fusion revolutionises how ergonomic factors are analysed, leading to more precise and efficient assessments for improving workplace design and human-tool interactions.

Traditionally, image processing grappled with challenges such as intra-class variations, scale changes, and viewpoint alterations [Den22]. Techniques like Histogram of Oriented Gradients (HOG) and Bag of Words (BoW) were limited in handling large datasets and had restricted learning capabilities. The landscape began to shift with the introduction of Convolutional Neural Networks (CNNs), pioneered by LeCun in 1989. This marked the beginning of a new era where the principles of biological neural networks were applied to computational problem-solving [Den22].

Like their biological counterparts, neural networks consist of interconnected nodes or 'neurons' that process and transmit information. This architecture enabled networks to learn from data, making decisions and predictions with increasing accuracy [JZ19]. CNNs emerged as game changers in this domain, excelling in tasks like image classification and object recognition due to their ability to learn spatial hierarchies of features. Other neural network types, such as Recurrent Neural Networks (RNNs) and Autoencoders, expanded the scope of applications to video frame analysis and image compression, respectively [JZ19].

The effectiveness of neural networks in image processing is primarily credited to fundamental techniques such as convolution, pooling, and backpropagation. Convolution, the process of applying filters to extract image features, is crucial for identifying patterns. Pooling reduces the spatial dimensions of the data, simplifying the amount of computation needed and minimising overfitting. At the same time, backpropagation is a method for fine-tuning the network's weights based on error rates, ensuring that the network learns effectively from its training data. This trio of techniques has shifted the paradigm from manual, computation-intensive feature extraction to automated, highly accurate processing, marking a significant advancement in the field of image processing [JZ19].

In contemporary applications, neural networks have had a transformative impact across a broad spectrum of image-processing tasks. They have enhanced image classification and significantly improved object detection and facial recognition technologies [JZ19, MJW+20]. This evolution has been pivotal in fields ranging from autonomous vehicles to security systems, where accurate and rapid visual data processing is crucial.

However, integrating neural networks into image processing also brings several challenges. One primary concern is the need for large datasets to train these networks effectively. Neural networks, particularly those designed for complex tasks, require vast amounts of data to learn and generalise effectively. This demand often poses logistical and ethical challenges, especially when collecting sensitive data such as facial images. Another significant challenge is the computational demands of these networks. The processing power required to train and run sophisticated neural network models is substantial, necessitating advanced hardware and optimised algorithms [JZ19, Den22]. This requirement can be a barrier, particularly in resource-limited settings.

Moreover, the interpretability of neural network outputs remains an area of concern. The 'black box' nature of these networks often makes it difficult to understand how they arrive at certain decisions or predictions [JZ19]. This lack of transparency can be problematic in applications where explainability is critical, such as medical diagnoses or legal scenarios.

In addition, as deep learning models become more complex, the need for novel models and parallel computing systems becomes more apparent. These advanced models and systems are essential for more effectively interpreting image content, especially in complex images or tasks requiring nuanced understanding. Developing these new models and systems is crucial for continuing to advance the capabilities of neural networks in image processing, ensuring they can meet the demands of increasingly sophisticated applications [Den22].

Looking ahead, the field of neural networks in image processing is poised for further advancements. The trend is shifting from large, complex networks to more efficient, lightweight models, focusing on the speed-accuracy balance. Deep learning algorithms are expected to evolve, leading to more capable networks. The increasing integration of AI in various industries, from healthcare to automotive, promises to drive further innovation in image processing [JZ19].

2.4.1 Human Pose Estimation

Human Pose Estimation (HPE), utilizing advanced 2D and 3D detection methods, has become an instrumental tool in ergonomics by providing detailed insights into human movement. This technology aids ergonomic research in understanding how people physically interact with their environments, enabling the design of spaces and tools better aligned with natural human motion. By accurately capturing and analyzing body postures, HPE directly informs ergonomic solutions, ensuring environments are more comfortable and less prone to causing injury, thereby bridging the gap between technological innovation and practical ergonomic application. The essence of HPE lies in estimating body parts or joints in photos and videos, which can be executed in either 3D or 2D formats, as seen in Figure 2.6. In 3D HPE, the task is to identify the 3D key points of the human body from an image or video, using x, y, and z coordinates, thereby computing the 3D posture from an RGB image. On the other hand, 2D HPE focuses on calculating the position of a person's joints in visuals like photos or videos [JB23].

One of the critical subdivisions of HPE is based on the number of persons in the image or video. Single-person HPE focuses on defining the human pose skeleton based on key points located for an individual. Techniques like Direct Regression, which maps the joints of the body or features of human body models, and Heatmap Regression, commonly used for 2D HPE and grouping key points location, are prevalent in single-person HPE [MJW+20].

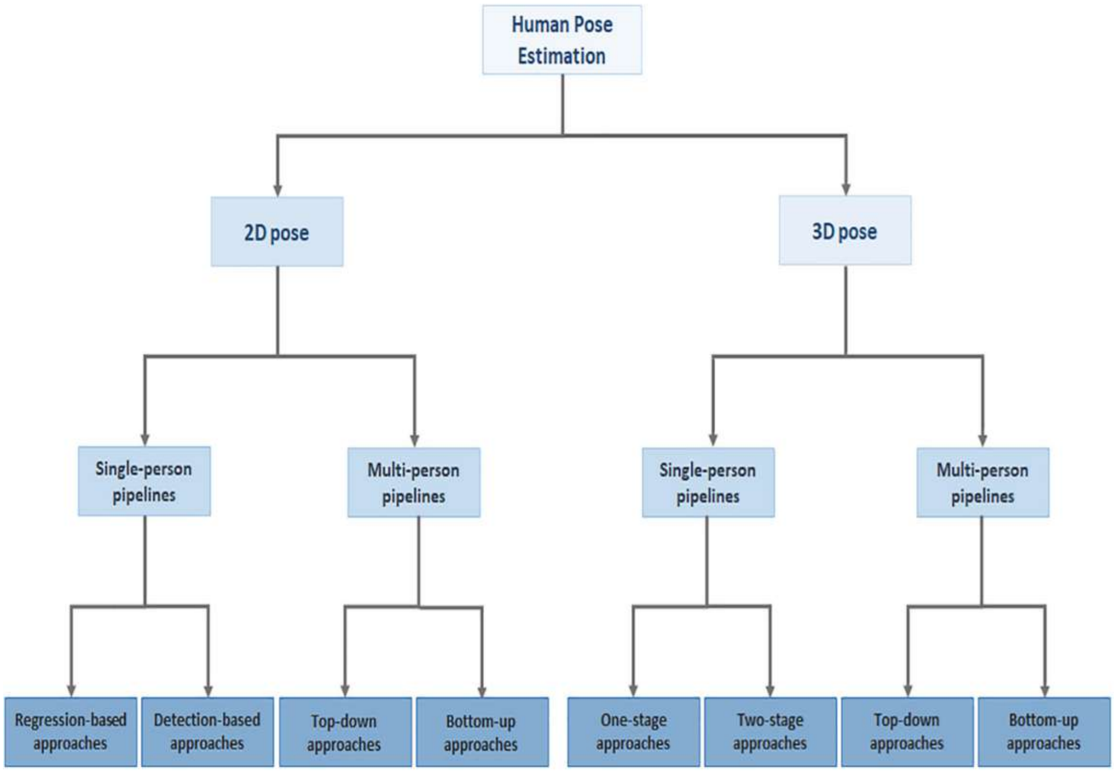


Figure 2.6: The taxonomy of human pose estimation [GA21]

In contrast, multi-person HPE is designed to identify the landmarks of multiple people. This becomes particularly challenging in crowded images where distinguishing individual poses requires sophisticated algorithms. Top-down and bottom-up approaches are typically employed in Multi-person HPE. The Top-Down method involves person detection followed by pose estimation, while the Bottom-Up approach works by computing key points for multiple persons simultaneously [JB23].

Another division of 3D Human Pose Estimation (HPE) is in the approach they take. In 3D HPE, one-stage methods directly infer the three-dimensional pose from images using deep neural networks (DNNs), focusing on capturing the intricate details of human anatomy without explicit 2D pose estimation. Techniques such as multitask learning and overcomplete autoencoders are employed to improve depth perception and joint interdependencies. Conversely, two-stage methods first estimate the pose in two dimensions using convolutional neural networks (CNNs) and then project this information into 3D space, often incorporating kinematic models to enhance the fidelity of the skeletal representation. These methods benefit from the rich spatial-temporal context provided by sequential image frames, with advanced models leveraging weakly-supervised and self-supervised learning to overcome the scarcity of labelled 3D data, thereby improving the robustness and accuracy of pose estimation across diverse environments [GA21].

Developing various deep-learning models has significantly enhanced the field of HPE. OpenPose, seen in 2.3, is known for detecting multiple individual human poses in real-time and identifying various key points on the body, including hands, feet, and facial landmarks [KSS+21]. Similarly, models like AlphaPose, Mask R-CNN or techniques like Iterative Error Feedback have contributed to the advancement of multi-person HPE with their unique approaches [JB23].

System	**Feature**
OpenPose	Real-time, multi-person 2D pose detection system with real-time performance
High-Resolution Net (HRNet)	Specializes in high-resolution representations with high-resolution output
DeepCut	Early multi-person pose estimation using deep learning, a pioneer in deep learning for pose estimation
AlphaPose	Regional Multi-Person Pose Estimation with improved accuracy
PoseNet	Lightweight model for real-time pose detection, optimized for real-time applications
Dense Pose	Maps human pixels in images to 3D body surface, converts 2D images to 3D body surfaces
TensorFlow Pose Estimation	Pose detection using TensorFlow utilizes TensorFlow framework
OpenPifPaf	Open-source, accuracy-focused pose estimation with emphasis on accuracy and open-source availability
YoloV8 Pose Estimation	Latest YOLO series adaptation for pose estimation adapts YOLO architecture for pose estimation

Table 2.3: Overview of Pose Estimation Systems

The primary difference between 2D and 3D HPE lies in their dimensionality. 2D HPE focuses on identifying and predicting body keypoint locations on a two-dimensional plane using X and Y coordinates. This approach does not take into account depth of the subject in the scene, making it more straightforward and less computationally intensive. However, lacking depth information can lead to ambiguities, especially in complex poses or when body parts overlap. In contrast, 3D HPE adds a Z-axis to the analysis, providing depth information. This depth perception is crucial for understanding the spatial positioning of body joints in three-dimensional space, making 3D HPE more suitable for applications where understanding the complete spatial orientation of the body is essential [GA21, MJW+20].

The complexity of the pose estimation task increases significantly when moving from 2D to 3D. 3D HPE requires more sophisticated algorithms and, often, more computational resources. This is partly because 3D models need to infer the depth information from inherently 2D images. This process often relies on large datasets with 3D annotations, which are more challenging to acquire than 2D datasets. This 3D HPE method typically requires datasets captured in controlled environments, like motion capture studios, to accurately annotate the depth information, limiting the training data's diversity and impacting the model's performance in real-world scenarios.

The choice between 2D and 3D HPE also depends on the application. 2D HPE is often sufficient for applications like simple activity recognition or user interface control, where the detailed spatial orientation of the body is not critical. However, for more complex applications such as advanced gesture recognition, virtual reality, or clinical gait analysis, 3D HPE is preferable because it provides a more complete understanding of human movement and posture. The paper by Bergman [GA21] emphasises that while 3D HPE presents more challenges, it offers richer information and potential for a wider range of applications, particularly those requiring a detailed understanding of human kinematics. Datasets play a pivotal role in the development and testing of HPE algorithms. For 2D HPE, datasets like MPII, MSCOCO, CROWDPOSE, and POSETRACK offer a wide range of images and videos for training and evaluating HPE models. These datasets vary regarding the number of images, key points annotated, and the scenarios they cover.

For 2D Human Pose Estimation, datasets typically consist of images or videos annotated with key points that mark significant body joints. These datasets vary widely in terms of the number of images, the diversity of poses, the complexity of the backgrounds, and the range of human activities they cover. Examples include the MPII Human Pose and COCO datasets, which are critical for training and benchmarking 2D pose estimation models. The MPII dataset, for instance, offers a variety of challenging poses and camera angles. In contrast, the COCO dataset provides a broad range of human poses and scenarios, including occlusions and different scales of human figures [JB23, GA21]. In 3D Human Pose Estimation, the datasets are more complex as they include depth information alongside the traditional 2D key points. These datasets are often generated in controlled environments using motion capture technology to capture the 3D position of body joints accurately. Examples include the Human3.6M dataset, one of the most extensive for 3D pose estimation, featuring millions of frames with detailed 3D annotations. These datasets are crucial for developing models that can understand and infer human bodies' spatial positioning and movement in three-dimensional space.

Overall, the availability and quality of these datasets significantly impact the progress and performance of HPE models. They provide the necessary data for training, testing, and refining algorithms, and their diversity ensures that the models are robust and effective across various real-world scenarios [GA21].

Human Pose Estimation (HPE) applications extend into various domains, demonstrating its versatility and importance in contemporary technology. In healthcare, HPE is instrumental in monitoring and guiding physical exercises, offering a valuable tool for physiotherapy and rehabilitation. It enables precise tracking of patients' movements, ensuring that exercises are performed correctly, thus aiding in faster and more effective recovery [JB23]. Additionally, HPE has become a fundamental tool in ergonomics. It is extensively used to analyse and improve workplace ergonomics, reducing the risk of work-related injuries. By assessing the posture and movements of employees in real-time, HPE helps design more ergonomic workspaces, contributing to better occupational health and productivity. In the field of man-machine interaction, HPE enhances human-computer interface systems. It allows for more intuitive and natural interactions between humans and machines, paving the way for advanced control systems in various applications, from interactive gaming to sophisticated robotic control. This technology fosters a more immersive and responsive experience in virtual environments, enhancing user engagement and effectiveness. Human Pose Estimation (HPE) plays a key role in the field of ergonomics within the production industry [PKKC22]. HPE provides invaluable insights into workplace practices by analysing and interpreting workers' movements and postures in real-time. This technology is instrumental in identifying motions and positions that may contribute to physical strain or repetitive stress injuries. As a result, it allows for the design of workstations and workflows that are more aligned with the physical capabilities and limitations of the workforce. Consequently, the application of HPE not only bolsters worker safety but also enhances overall productivity and the quality of the products, as ergonomic improvements often lead to a more efficient and comfortable working environment.

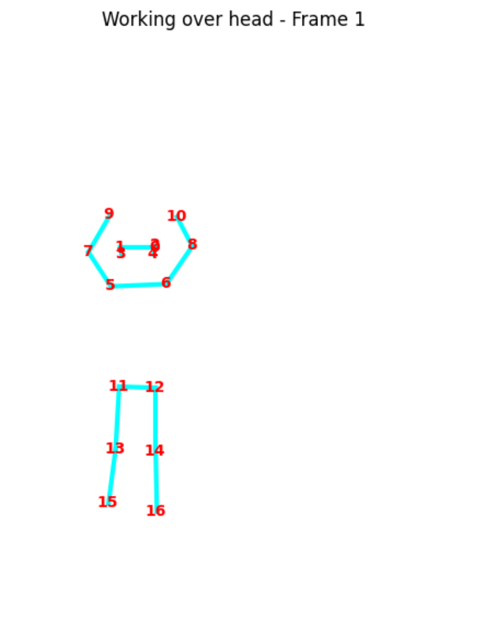


Figure 2.7: Human Pose Estimation

Index	Key point
0	Nose
1	Left-eye
2	Right-eye
3	Left-ear
4	Right-ear
5	Left-shoulder
6	Right-shoulder
7	Left-elbow
8	Right-elbow
9	Left-wrist
10	Right-wrist
11	Left-hip
12	Right-hip
13	Left-knee
14	Right-knee
15	Left-ankle
16	Right-ankle

Figure 2.8: Key points index

Despite the remarkable progress in HPE, challenges persist, particularly in crowded scenes and occlusion, where detecting individual poses becomes increasingly complex [JB23]. The future of HPE holds promising potential, with the possibility of moving beyond 2D and 3D to analyze high-dimensional datasets more effectively.

In conclusion, the evolution of human pose estimation, marked by the transition from classical approaches to sophisticated deep learning techniques, underscores the dynamic nature of this research area. With ongoing advancements, HPE continues to broaden its horizons, offering innovative solutions and applications that resonate across numerous domains [MJW+20, JB23].

2.5 Neural Networks for Optimisation and Transfer

In the domain of ergonomics, the advent of neural network technology presents an innovative path for enhancing human well-being. Specifically, applying this technology in ergonomics involves processing video footage to analyze human postures. By employing neural optimization combined with neural style transfer and generative adversarial networks (GANs), it becomes feasible to identify and rectify suboptimal ergonomic postures. The core concept is inspired by neural style transfer, which traditionally blends one image's content with another's style. In this context, the 'content' is the original posture captured in the video, often depicting an unergonomic stance, while the 'style' is represented by an ergonomically correct posture.

The objective is to optimize the Rapid Upper Limb Assessment (RULA) scores, described in Chapter 2.3, through neural network optimization. In applying neural networks to corrective ergonomics, the challenge lies in balancing the optimization of postures with the preservation of the original intent of the action. For instance, if the original action involves lifting, the optimized posture will not default to standing but instead suggest a more ergonomic way of lifting. The integration of GANs, particularly the use of a discriminator network, aids in ensuring that the suggested postures are realistic and human-like. This innovative approach to corrective ergonomics through neural networks could significantly enhance ergonomic awareness among non-ergonomic experts.

2.5.1 Neural Style Transfer: An Overview

Neural style transfer (NST), primarily known for its applications in art and technology, finds a unique and impactful application in the field of image processing. It represents an intersection of art and technology, merging the artistic allure of traditional painting with the advancements of modern computer science and neural networks. NSTs can take the style of an image and apply it to the content of another as seen in Figure 2.10. Historically, painting has been a revered form of art, captivating human interest for millennia with masterpieces like Van Gogh's "The Starry Night". Traditionally, recreating these styles required skilled artists and significant time. However, since the mid-1990s, there has been a growing interest among computer science researchers in automating the process of turning images into synthetic artworks, mainly through non-photorealistic rendering (NPR) [JYF+19].

While NPR techniques made significant strides in computer graphics, they often focused on specific artistic styles and struggled with flexibility and extending to others. The approach to style transfer in computer vision was more general, focusing on texture synthesis - transferring textures from one image to another. However, these methods had limitations, primarily relying on low-level image features and often failing to capture the essence of image structures effectively [JYF+19].

The landscape of NST underwent a significant transformation with the innovative approach introduced by Gatys et al. in their 2015 paper [GEB15]. While Convolutional Neural Networks (CNNs) were already established, Gatys and his team pioneered their application in NST. Their noval work demonstrated that CNNs could effectively extract content from any photograph and style information from famous artworks. This novel approach uses CNN feature activations to merge a given photo's content with the artwork style. The process entails iteratively optimizing an image to match the desired CNN feature distributions, capturing both the photo's content and the artwork's style. This technique has been remarkably successful in producing images that not only mirror the appearance of the chosen artwork but also retain the essence of the original photo's content [GEB15].

Gatys et al.'s work created visually stunning images and launched NST as a new field, attracting extensive attention in academia and industry. Follow-up studies sought to improve or extend the NST algorithm, leading to various applications in the industry, such as Prisma, Ostagram, and Deep Forger. Despite this progress, as of 2018, no comprehensive survey was summarizing these advances and challenges in NST [GEB15].

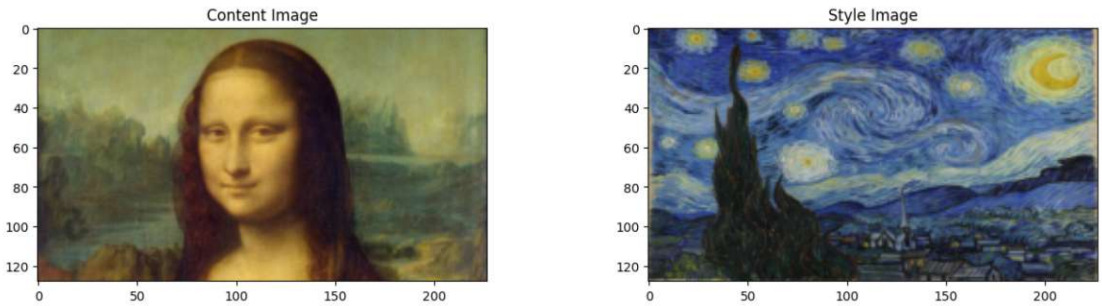


Figure 2.9: Content and Style images used for demonstrating NST

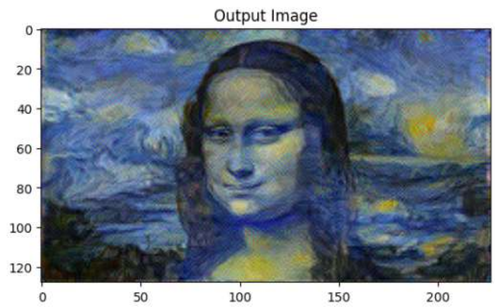


Figure 2.10: Output of NST [Author Created]

The NST process begins with the selection of two images. The 'content image' is the primary picture whose content is to be preserved, and the 'style image' is usually an artistic work whose style characteristics are to be transferred. A pre-trained CNN, such as the VGG network, analyses these images. These networks are well-suited for this task due to their ability to extract and interpret features at various levels of complexity. Feature extraction is a pivotal part of NST. When passed through the CNN, the content image undergoes a transformation where its features are captured at different layers. Deeper layers, responsible for higher-level content features such as objects and overall structure, are generally preferred for content representation. The CNN processes the style image, emphasizing capturing its stylistic elements like textures, colours, and brushstrokes. These elements are represented across various layers within the CNN [GEB15].

The core of NST lies in defining and minimizing specific loss functions. The 'content loss' measures the deviation in content between the transformed and original content images. Lower content loss indicates higher retention of original content. The 'style loss' evaluates the difference in style between the transformed image and the style image, often using statistical measures from CNN's feature maps. Sometimes, a 'total v qariation loss' is included to ensure spatial coherence in the final image.

The optimization process in NST starts with a target image, which could be the content image, a blank canvas, or an arbitrary image. This target image is iteratively modified to minimize the combined loss. The adjustments aim to make the target image's content more akin to the content image while aligning its style with that of the style image. This iterative process uses backpropagation, akin to the training of neural networks [GEB15]. The result of this iterative optimization is an image that harmoniously blends the content of the content image with the style of the style image. This synthesis demonstrates the potential of NST not only in artistic fields but also in various applications such as advertising and entertainment.

The emergence of NST marked a departure from earlier artistic rendering methods that did not utilize CNNs—prior to NST, artistic stylization research expanded into NPR, focusing on image-based artistic rendering (IB-AR). These techniques included stroke-based rendering, region-based, example-based rendering, and image processing and filtering.

However, they often needed more flexibility, style diversity, and effective image structure extraction [JYF+19]. NST's development can be traced back to visual texture modelling and image reconstruction techniques. Visual texture modelling, the heart of texture synthesis, involves parametric texture modelling with summary statistics and non-parametric modelling with Markov Random Fields (MRFs). NST particularly benefited from the parametric approach, which captures image statistics from sample textures using summary statistical properties [JYF+19].

In conclusion, neural style transfer represents a significant leap in the synthesis of art and technology, providing a versatile tool for creating synthetic artworks with a broad range of styles. It stands as a testament to the evolving capabilities of neural networks and their potential to revolutionize how we perceive and create art.

2.5.2 Gradient Descent and Optimization

Gradient descent is a crucial algorithm in machine learning and optimization, widely employed to minimize an objective or loss function. This process is central to many applications within statistical modelling and deep learning. The algorithm operates by iteratively adjusting the function's parameters, denoted as θ, in small increments. These adjustments are explicitly made in the direction opposite to the gradient of the objective function at the current point, where θ represents model parameters such as weights in neural networks. The objective of gradient descent is to identify parameter values that minimize the loss function, thereby enhancing the model's performance on given tasks [Rud16].

The objective function, often a loss function in machine learning contexts, quantifies the model's prediction error. For instance, a joint objective function in neural networks is the mean squared error, measuring the disparity between predicted and actual outputs.

The gradient, a vector comprising partial derivatives of the objective function concerning each parameter, intuitively represents the function's slope in multidimensional space. It points towards the steepest ascent direction; thus, moving opposite to the gradient direction leads to the steepest descent, which is the foundational concept behind GD.

$$\theta = \theta - \eta \cdot \nabla J(\theta) \tag{2.3}$$

Parameters undergo iterative updates following the rule 2.3, where η denotes the learning rate. This learning rate is a critical hyperparameter determining the step size during descent. A large learning rate might overshoot the minimum, whereas a small rate may result in slow convergence.

Convergence in GD is achieved when there is a negligible decrease in the objective function, indicating that a local minimum, or ideally a global minimum, has been reached. However, challenges arise in complex, non-convex landscapes typical of deep learning, where the algorithm might get stuck in local minima or saddle points [Rud16]. GD manifests in several variants: Batch Gradient Descent computes the gradient using the entire dataset at each step. Stochastic Gradient Descent (SGD) estimates the gradient using a single data sample per iteration. Mini-batch Gradient Descent, a balance between the former two, uses a dataset subset for gradient computation. Critical considerations for effective GD implementation include feature scaling, learning rate scheduling, and the choice of initialization, each influencing the optimization's efficiency [SCZZ19] [6]. GD's applications are widespread in machine learning:

Gradient descent is a fundamental optimization method across various machine learning applications, enabling backpropagation in neural networks, facilitating the derivation of

[6] https://www.geeksforgeeks.org/gradient-descent-algorithm-and-its-variants/

the best-fit line in linear regression, assisting binary classification in logistic regression, and optimizing class margins in Support Vector Machines (SVMs).

Despite its extensive applications, GD encounters issues like entrapment in local minima, especially in non-convex functions. Strategies such as momentum introduction or varied initialization techniques can help mitigate these challenges. Furthermore, selecting appropriate hyperparameters like learning rate and batch size is crucial, often necessitating experimentation.

In conclusion, while GD is a versatile and potent tool in machine learning for model optimization, its effectiveness relies heavily on judicious hyperparameter selection and a deep understanding of the model's and data's intricacies.

Backpropagation and gradient descent are interconnected in neural network training. Backpropagation efficiently computes the gradients of the cost function relative to each weight and bias, revealing how much each neuron contributes to output errors. These gradients are essential for gradient descent, which uses them to adjust the network's weights and biases (seen in Figure 2.11) iteratively. This synergy ensures that changes in the network minimize the cost function, facilitating learning from errors and improving performance. Thus, while backpropagation determines the direction of the steepest descent, gradient descent manages the step size in that direction, harmonizing error correction with network optimization.

Backpropagation is a robust algorithm used in neural networks to efficiently compute the gradients of the cost function with respect to the weights and biases, a process fundamental to the learning capability of the network.

This algorithm applies the chain rule of calculus, which allows for the computation of complex derivatives through the multiplication of simpler ones. At the heart of backpropagation is the computation of the error 'δ' at each neuron, which essentially measures how much that neuron contributed to the overall error in the output [Nie15]. Starting from the output layer, backpropagation works backwards through the network, using the chain rule to calculate the error gradient for each layer. This is done by first determining the error in the output layer ('δ'L) and then iteratively propagating this error backwards, layer by layer, to compute the errors in all previous layers [3]. The algorithm's efficiency stems from its ability to reuse computed values, thus avoiding redundant calculations that would be necessary if each partial derivative of the cost function was computed independently. Consequently, backpropagation provides a computationally efficient method to adjust the weights and biases in a direction that minimizes the cost function, enabling neural networks to learn from their errors and improve performance [Rud16]. This systematic and efficient approach to gradient computation not only enhances the learning speed but also deepens our understanding of how changes in weights and biases affect the overall behaviour of the network [SCZZ19].

In neural style transfer, the role of gradient descent is pivotal for the optimization process. Fundamentally, NST involves blending the style of one image with the content of another through deep neural networks. This task requires precisely aligning features extracted

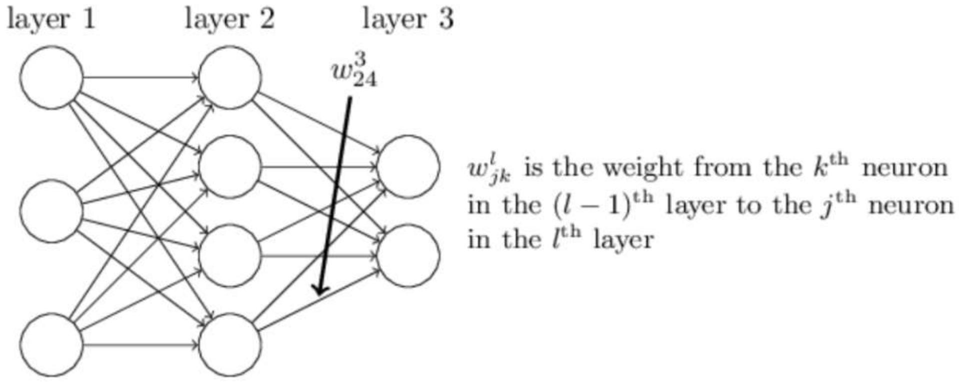


Figure 2.11: Neural Network Scheme [Nie15]

from the style and content images. This alignment is achieved through a loss function that encompasses both style and content loss. Gradient descent plays a crucial role here, as it is used to adjust the image to minimize this combined loss iteratively. Starting with an initial image – either a copy of the content image or a white noise image – gradient descent iteratively tweaks the pixel values to reduce the difference in content and style features relative to the target images. This process involves calculating the gradient of the loss function with respect to the image pixels and adjusting the pixels to reduce the loss, thereby gradually transforming the image into a style-transferred version that retains the core of the content while adopting the artistic nuances of the style reference.

The application of gradient descent in neural style transfer presents unique challenges, particularly in navigating the complex, high-dimensional landscape of the image space. Since the aim is to find a visually pleasing balance between the content and style of the two input images, the optimization process can be sensitive to the choice of hyperparameters, such as the learning rate and the relative weighting of the style and content losses. A high learning rate may result in erratic changes that converge to a satisfactory solution, whereas a low rate may lead to slow progress and potentially getting stuck in local minima. Additionally, the non-convex nature of the problem means that gradient descent might converge to different solutions depending on the initialization and specific parameter settings, making repeatability a challenge. To address these issues, NST algorithms often employ advanced variants of gradient descent, such as Adam or RMSprop, which are more adept at handling noisy gradients and adaptive learning rates, thus enhancing the stability and efficiency of the style transfer process.

This section provides an overview of the primary optimization algorithms that are used in training deep neural networks, detailing their roles and the advancements they offer in managing the complexities associated with neural network training.

Stochastic Gradient Descent (SGD) updates model parameters utilizing a single training example at a time. This method enhances convergence compared to batch gradient descent by introducing noise through its updates, which can aid in escaping local minima. However, the induced noise can also lead to variability in the training process, making it less stable than other methods. SGD is particularly effective in large-scale data scenarios and remains a foundational tool in neural network optimization [Shu23].

Adagrad improves upon SGD by adapting the learning rates for each parameter independently, based on the historical sum of squared gradients. This adaptation is particularly beneficial for handling sparse data, as it allows infrequent features to be updated more aggressively than frequent ones. Despite its potential to significantly enhance the robustness of SGD, Adagrad may suffer from a diminishing learning rate over extended training periods, potentially halting learning prematurely [Shu23].

RMSprop addresses the diminishing learning rate issue observed in Adagrad by modifying the learning rate based on a moving average of squared gradients. This method helps maintain an adequate learning rate throughout the training process, facilitating more consistent updates and improved convergence in practice [Shu23].

These optimization algorithms are important for the effective training of deep neural networks. Each method offers advantages and is suited to different types of data and network structures. The choice of algorithm can significantly impact the performance and efficiency of model training. While these methods are widely utilized and well-regarded, ongoing research continues to refine their effectiveness and develop new strategies to address the challenges presented by increasingly complex neural network architectures [Shu23, Sun19].

2.5.3 Generative Adversarial Network GAN

Generative adversarial networks are a class of artificial intelligence algorithms used in unsupervised machine learning, introduced by Ian Goodfellow and his colleagues in 2014 [GPAM+20]. GANs consist of two key components: a generator (G) and a discriminator (D), essentially two neural networks competing against each other. The generator's role is to create indistinguishable data from actual data, while the discriminator's task is to distinguish between genuine and artificially generated data. They operate in a minimax game framework, as seen in Formula 2.4, where the generator tries to maximise the probability of the discriminator making an error, and the discriminator tries to minimise this probability. This competitive process generates highly realistic data, with applications ranging from image and video generation to complex problem-solving in various domains. The unique architecture of GANs enables them to learn and replicate intricate data distributions, making them a powerful tool in deep learning and artificial intelligence.

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{zsimp_z(z)}[\log(1 - D(G(z)))] \tag{2.4}$$

During each training cycle, the generator attempts to produce data that the discriminator evaluates, ranging from images, text, video or sequences of key points representing human poses. Over time, this leads to the development of a model capable of generating a variety of realistic, data representation [GPAM+20].

During the initial training stages, the output from the generator in generative adversarial networks may significantly deviate from realistic outputs, posing challenges in effective gradient learning. This issue is particularly pronounced when the generator begins to learn from a complex or diverse dataset. To address this, the generator's training can be optimised to maximise the likelihood of the discriminator making classification errors. This strategy enhances the learning gradients by providing stronger feedback, which is crucial for improving the generator's output quality in the early training stages.

Additional challenges identified in the literature [ITDÁ23] include the instability of GAN training, the occurrence of mode collapse, and the complexities involved in evaluating the quality of generated data. These problems are not exclusive to any single application domain but are shared across various fields where GANs are employed to synthesise realistic data from complex distributions.

The instability refers to the delicate balance needed between the generator and the discriminator during training; if either becomes too powerful too quickly, it can lead to suboptimal learning. Mode collapse occurs when the generator produces a limited variety of outputs, restricting the model's ability to generate a range of data distributions. Addressing these challenges requires careful tuning of the model architecture and training parameters. Moreover, evaluating the quality of the generated postures is a complex task. Traditional metrics used in GAN training might not adequately capture the nuances of distribution of data, necessitating the development of specialised assessment criteria tailored for ergonomic evaluation. By addressing these challenges, GANs can be more effectively used to create a wide range of realistic data distributions suitable for various applications. The theoretical foundation of GANs is deeply rooted in their capacity to mimic and replicate complex data distributions as outlined in the original generative adversarial networks paper [GPAM+20].

In a standard GAN setup, the generator (G) creates samples from noise (z) (drawn from a noise distribution $p_z(z)$), and the discriminator (D) evaluates samples from both the generator and the actual data distribution. The generator aims to produce data that the discriminator cannot easily classify as fake. This process is formalised in a minimax objective function:

Here, $\mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)]$ represents the expectation of the discriminator correctly identifying real data. In contrast, $\mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))]$ is the expectation of the discriminator identifying fake data generated by G. The discriminator tries to maximise this function, correctly distinguishing between real and fake data, while the generator tries to minimise it, aiming to fool the discriminator [GPAM+20].

In conclusion, generative adversarial networks represent a significant advancement in unsupervised machine learning. By leveraging a competitive minimax game between two neural networks—a generator and a discriminator—GANs excel at generating highly realistic and intricate data across various domains. Despite challenges such as training instability, mode collapse, and the complexity of evaluating generated data quality, the foundational principles and continuous improvements in GAN technology hold substantial promise for enhancing the capabilities and applications of artificial intelligence.

CHAPTER 3

State of the Art

This chapter examines examines prior research on AI in ergonomics. Key focus areas include neural networks and their potential in ergonomic risk identification and correction. This synthesis of technology and ergonomics forms the basis for proposed AI-driven solutions to improve workplace health and safety.

The chapter summarizes research papers, categorized into three sections: 3.1 Overview of Ergonomic Assessment Techniques, 3.2 Automated Ergonomics Assessment, and 3.3 Generative Approaches and Optimization in Human Pose Estimation (HPE). Each section, as seen in Table 3.1 researches its area, offering insights into the latest methodologies and discoveries. The first section focuses on the variety of techniques used for ergonomic assessment, encompassing both traditional and new approaches. The second section explores the field of automated systems in ergonomics, highlighting how technology is revolutionizing the way ergonomic analysis is conducted. Lastly, the third section discusses generative methods and optimization strategies in HPE, emphasizing their impact on improving human-machine interaction and workplace safety. This comprehensive review aims to provide an integrated understanding of these interconnected fields, thereby laying a robust foundation for a novel approach for AI-based ergonomic corrections.

3.1 Overview of Ergonomic Assessment Techniques

This section explores the landscape of ergonomic assessment in industrial environments. The focus is on two distinct aspects: the comparative analysis of different ergonomic assessment methods and the general overview of the integration of advanced technologies, such as machine learning, in ergonomic practices. This discussion sets the stage for a deeper understanding of current trends and future directions in industrial ergonomics, particularly in the context of manufacturing and the challenges posed by Industry 4.0.

Group Name	Papers
Overview of Ergonomic Assessment Techniques	- Machine learning in manufacturing ergonomics [LLRL21b] - Different approaches of conducting ergonomic assessment utilizing digital human models and motion capture in industrial site assembly [FRS23]
Automated Ergonomics Assessment	- Spatio-Temporal Pyramid Graph Convolutions for Human Action Recognition and Postural Assessment [PSH+19] - A Multi-Task Learning Approach for Human Activity Segmentation and Ergonomics Risk Assessment) [PB21] - Applying incremental Deep Neural Networks-based posture recognition model for ergonomics risk assessment in construction[ZO21] - Development of a fully automated RULA assessment system based on computer vision [NK21] - Applying incremental Deep Neural Networks-based posture recognition model for ergonomics risk assessment in construction [PBL24] - ErgoMaps: Towards Interpretable and Accessible Automated Ergonomic Analysis [KWS22] - Combining ergonomic risk assessment (RULA) with inertial motion capture technology in dentistry—using the benefits from two worlds. 2[MGHF+21] - Ergonomic postural assessment using a new open-source human pose estimation technology (openpose) [KSS+21] - Ergoexplorer: Interactive ergonomic risk assessment from video collections [FRM+22] - Industrial ergonomics risk analysis based on 3d-human pose estimation [PKKC22]
Generative approaches & optimization in HPE	- A virtual element-based postural optimization method for improved ergonomics during human-robot collaboration. [EMMV+22] - Learning view-invariant gait features with Two-Stream GAN [WSH+19] - Frame-GAN: Increasing the frame rate of gait videos with generative adversarial networks [XAS+20] - Appearance and Pose-Conditioned Human Image Generation using Deformable GANs [SSLS19]

Table 3.1: Classification of Papers into Ergonomic Research Groups

The paper by Clara Fischer et al. [FRS23] provides a comparative analysis of three different ergonomic assessment methods applied in an industrial assembly setting. The study focuses on evaluating and contrasting these methods to determine their effectiveness in identifying ergonomic risks in a workplace scenario.

The experiment conducted involved a practical assembly task set up at the TU Wien Pilot Factory. The task included constructing a tower from various components, which required a range of physical movements and postures. The methods compared for ergonomic assessment were:

1. **Manual Evaluation:** This traditional approach uses paper and pencil to assess physical movements based on visual observations. While straightforward, it relies heavily on the evaluator's expertise and judgment.
2. **Automated Evaluation with a Digital Human Model:** This method uses software to simulate human movements and objectively assess ergonomic risks. It provides detailed insights, but it requires specific technical resources and expertise.
3. **Motion Capture:** This technique involves recording the physical human movement using motion capture technology, which is then applied to a digital human model for ergonomic evaluation.

The results from these methods varied, highlighting the strengths and weaknesses of each approach. The manual method was found to potentially overestimate ergonomic risks due to the subjective nature of visual assessment. In contrast, while more precise, the automated digital human model and motion capture methods could underestimate certain risks if not calibrated or utilized with sufficient detail. The study suggests that while automated methods can offer more objective and detailed assessments, they must be carefully calibrated and implemented to avoid underestimating ergonomic risks. The findings underscore the importance of choosing the proper ergonomic assessment method that balances accuracy, practicality, and resource availability, especially in complex industrial environments.

Modern advancements in AI could remedy the limitations of traditional motion capture techniques. In the paper by Lee et al. [LLRL21b], Lee et al. emphasize the need for integrating machine learning techniques into ergonomics research within the manufacturing sector. This integration is important to address the challenges of Industry 4.0, characterized by advancements in robotics and AI. Lee et al. suggest that traditional ergonomics methods should evolve to incorporate ML, enhancing their effectiveness and breadth. The paper proposes specific research areas, such as integrated worker-specific risk analysis, which combines physical and mental workloads in ergonomics models. It also highlights the potential of ML in predicting and managing ergonomics risks in manufacturing, especially when considering individual operator characteristics and an ageing workforce.

Furthermore, the paper discusses the significant role of human-machine interaction in modern manufacturing systems. It underscores the importance of developing ergonomic principles and standards for human-robot cooperation, considering the varying cognitive features of operators. The use of ML in these contexts can facilitate the reduction of operator workload and enhance safety. The paper also explores the concept of optimal decision support, where ML can aid in early risk detection and provide real-time, optimal decision-making to prevent injuries and production losses. This approach includes evaluating ergonomics from various angles, such as staff performance, health, safety, and environmental impact, and integrating these factors with intelligent equipment and production processes. The paper highlights the necessity for collaborative efforts across multiple disciplines to further research and application of ML in manufacturing ergonomics.

3.2 Automated Ergonomic Assesment

The integration of artificial intelligence (AI) into ergonomic risk assessment represents a transformative approach to enhancing workplace safety and efficiency. Ergonomics, fundamentally concerned with optimizing human well-being, significantly benefits from AI's analytical and predictive capabilities and data-driven insights. Automated ergonomic assessment, facilitated by AI, involves precise techniques such as pose estimation, motion tracking, and predictive analytics to identify potential risks more accurately and swiftly than traditional methods. Specifically, pose estimation technology enables the real-time analysis of workers' postures, identifying deviations from optimal ergonomic positions that could lead to injuries. This data, coupled with AI's capacity for large-scale data processing, allows for proactively predicting risk areas, enabling interventions before injuries occur.

In the paper by Nayak et al. [NK21], an approach that leverages the capabilities of human pose estimation to automate the Rapid Upper Limb Assessment (RULA) process. This method focuses on detailed aspects like wrist positioning and arm movement. The significance of this paper lies in its ability to transform ergonomic assessments from a manual, time-consuming process into an efficient, accurate, and automated system. This advancement streamlines the process and greatly reduces the likelihood of human error, ensuring more consistent and reliable results. By employing this technology, workplaces can proactively address ergonomic risks, enhancing worker safety and well-being.

Next, Parsa et al. [PSH+19] paper is a creative contribution to the field of computer vision, particularly in its application to ergonomic risk assessment in occupational safety. This work addresses the complex challenge of recognizing human actions and interactions with objects and the environment, a task of critical importance due to its wide-ranging applications across various domains such as production, surveillance, health monitoring, and occupational safety. The significance of this paper lies in its approach to action recognition through the development of the Spatio-Temporal Pyramid Graph Convolutional Networks (ST-PGN). This method integrates skeletal structure knowledge, hierarchical joint relationships, and a data-driven framework, which collectively enhance the precision of ergonomic risk assessments. By accurately modeling and predicting human movements and interactions with objects, this approach helps identify potential ergonomic issues, such as improper lifting techniques or awkward postures. This algorithm is adept at learning transitions between actions and is well-suited for real-time applications, a key advantage in dynamic and unpredictable environments. It is designed to tackle inherent challenges in action recognition, such as dealing with cluttered backgrounds, occlusions, and viewpoint variations, which are common in dynamic and complex environments like manufacturing assembly lines. Compared to state-of-the-art algorithms such as Spatio-Temporal Graph Convolutional Network (ST-GCN), the ST-PGN stands out due to its fewer graph convolution kernels, simplifying the model without compromising performance. This efficiency is particularly beneficial in real-time settings with critical computational resources and response time.

The versatility of ST-PGN enables it to generalize across various environments, making it a robust solution for real-time ergonomic risk assessment. The model's feature pyramid architecture autonomously captures the correlation between body parts, eliminating the need for hand-coding body-part relations and enhancing adaptability. What sets this research apart is its utilization of graph convolutional networks that extract features from the entire hierarchy of the skeleton. This method demonstrates enhanced performance, outperforming state-of-the-art action recognition algorithms in public benchmark datasets. Furthermore, incorporating online action recognition techniques into the model increases its ability to conduct postural assessments more effectively.

Finally, integrating the ST-PGN algorithm with a traditional ergonomic risk index (REBA) showcases its practical application and potential value in assessing musculoskeletal disorders, a significant concern in occupational safety. The research's ability to correlate action with time-varying posture and ergonomic risk in real-time represents a progress in the domain of ergonomic risk assessment.

Further research by Parsa et al. [PB21] introduces an innovative approach combining Human Activity Evaluation (HAE) and Human Activity Recognition (HAR). By building upon the foundational work detailed in earlier research, the study represents an advancement in applying computer vision and deep learning methodologies to the complex challenge of assessing human posture and activities in industrial and workplace settings.

The core innovation of this research is its multi-task learning framework, which synergistically combines the tasks of activity segmentation and ergonomic risk assessment, as seen in Figure 3.1. This integrated approach represents a major leap in the accuracy and efficacy of postural evaluations, addressing the nuances and variations in human activities that significantly impact ergonomic risk. The framework's novelty lies in its ability to leverage the intricate dynamics of human movement, using a combination of Graph Convolutional Networks (GCN) and Encoder-Decoder Temporal Convolutional Networks (ED-TCN), further enhanced with Long Short-Term Memory (LSTM) networks for refined activity assessment.

The Paper by Zhao and Obonyo [ZO21] developed Convolutional Long Short-Term Memory (CLN) network model is novel for its ability to incrementally learn and adapt to new postures without forgetting previously learned ones. Before, models could adapt to new information rapidly, but that led to inadvertently impaired memory of previously learned tasks, leading to a phenomenon known as "Catastrophic Forgetting".This is crucial

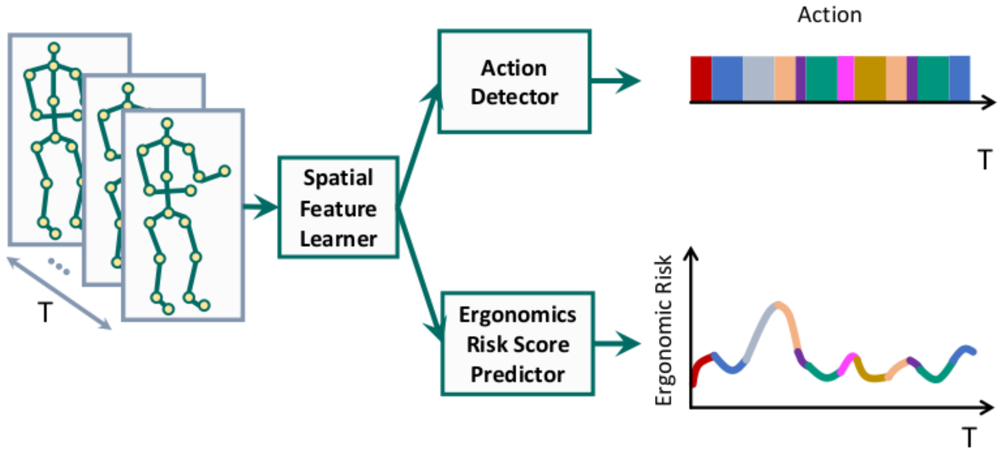


Figure 3.1: Multi-task activity segmentation and ergonomics risk assessment pipeline [PB21]

in the dynamic and varied environment of construction sites, where workers engage in various activities and postures. The model's architecture combines convolutional neural networks' spatial feature extraction capabilities with the temporal sequence learning strengths of Long Short-Term Memory (LSTM) networks. This hybrid design allows the model to recognize and classify complex worker postures over time more accurately. The application of this model in ergonomics risk assessment is particularly valuable. The model can provide critical insights for proactive musculoskeletal disorder (MSD) prevention by accurately identifying and assessing workers' postures.

The paper by Pistolesi et al. [PBL24] introduces a novel approach to ergonomic risk assessment in industrial environments. This research stands out for integrating a smart-watch and LiDAR (Light Detection and Ranging) technology to monitor and assess worker postures in real-time, aligning with the human-centric approach of Industry 4.0. The system's design is noteworthy for its emphasis on privacy and non-intrusiveness, using data that doesn't reveal the worker's identity. The methodology successfully addresses the ergonomic risk assessment challenge in a manufacturing setting, paving the way for safer and more efficient work practices. The exploration into applied and secondary research shaping the future of ergonomic technology goes beyond basic principles, delving into how initial theories and methodologies are being advanced, refined, and adapted for more complex and nuanced real-world applications. This area of study emphasizes the enhancement of practical applicability and interpretability, demonstrating innovative applications of foundational ergonomic research to create sophisticated and impactful solutions in various settings.

"The paper by Kostolani et al. [KWS22] addresses the limitations of ergonomic assessment using pose estimation, which often fall short of providing interpretable results for end-users. ErgoMaps bridges this gap by converting complex ergonomic data into intuitive and accessible heatmaps. These heatmaps facilitate the visual identification of high-risk areas on a shop floor, with zones of greater ergonomic risk highlighted in red as seen in Figure 3.2. This enhancement makes specific ergonomic concerns easier to identify and understand and enables the practical application of AI-driven ergonomic assessments in enhancing workplace safety measures.

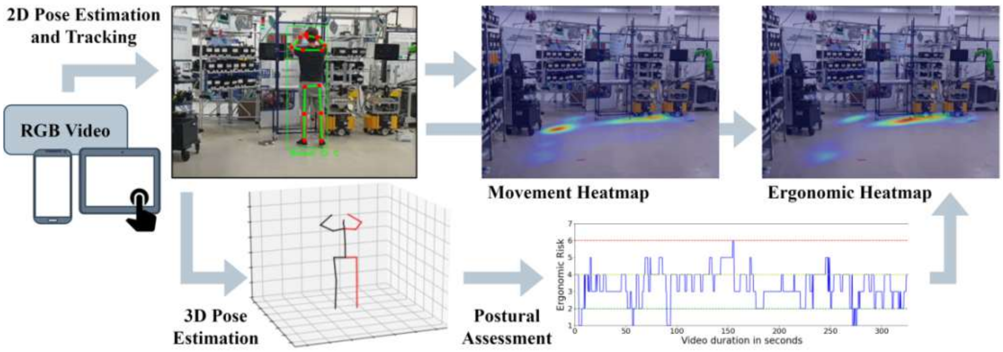


Figure 3.2: Workflow and output of the ErgoMaps method [KWS22]

A notable application of the output heatmaps is analysing shop floor layouts and workstations. By clearly indicating areas of high ergonomic risk, ErgoMaps empowers managers and safety professionals to redesign workspaces for improved ergonomics. This proactive approach can lead to more effective interventions and optimisations in workplace design, significantly advancing occupational health and ergonomics. The ability to visually map and understand ergonomic risks provides a powerful tool for making data-driven decisions, ultimately contributing to healthier, more efficient, and safer work environments. The introduction of explainability into the assessment of ergonomic scores addresses a critical gap often found in current ergonomic research, enhancing the usability and application of these assessments in real-world settings. This advancement in explainability supports more informed decision-making and fosters greater transparency and accountability in ergonomic practices.

In the paper by Maurer-Grubinger [MGHF+21], the researcher innovatively integrated RULA with motion capture to evaluate ergonomic risks in dentistry. This methodology, spanning five complexity levels, transforms raw inertial sensor data into a structured analysis ranging from basic RULA scores to detailed relative angle distributions. This method starts with the RULA score, providing an initial ergonomic load assessment. It then progresses to a more detailed relative RULA score distribution, illustrating the frequency of different scores. The third step involves the RULA steps score, focusing on specific body parts. The fourth level further refines this by analyzing the relative occurrences of these step scores. Finally, the most detailed level involves relative angle distributions, providing granular data about specific joint movements. This structured approach, from a broad overview to a detailed joint analysis, offers comprehensive insights into ergonomic risks in dentistry. This gradation allows for precise identification of ergonomic risks, enhancing the assessment's effectiveness.

The study's significance lies in its comprehensive approach, enabling detailed examinations of work-related postures and movements. By providing a nuanced and multi-layered analysis, the methodology addresses the specific ergonomic challenges in dentistry, offering insights that could improve occupational health and workplace design.

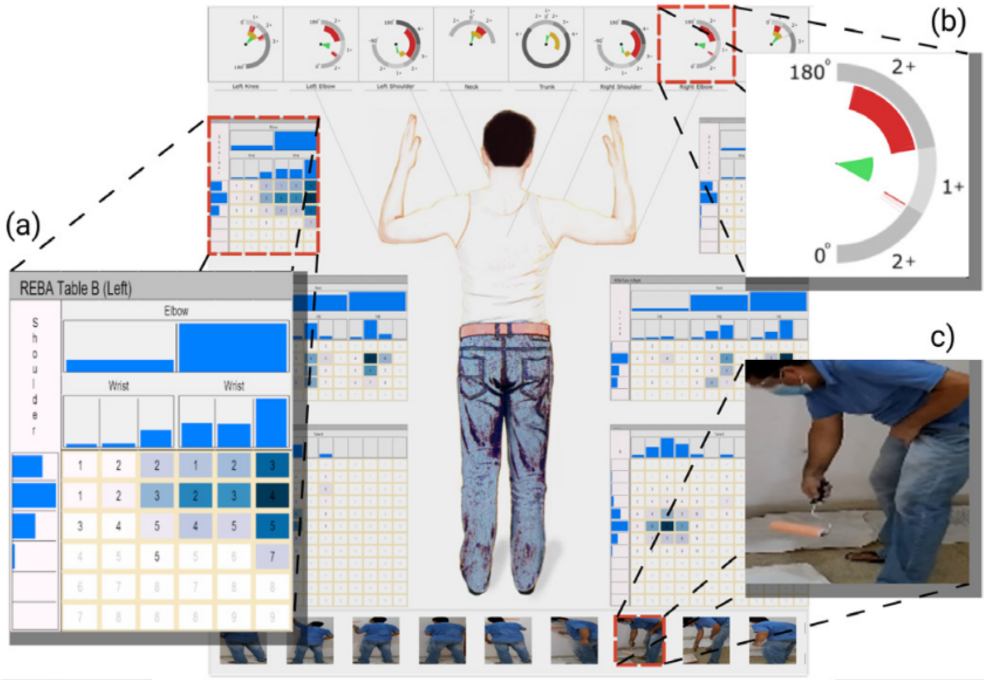


Figure 3.3: ErgoView: Table B from the REBA assessment is utilized to display the scores for the wrist, lower arm, and upper arm on the left side of the body. Table B from the REBA assessment is utilized to display the scores for the wrist, lower arm, and upper arm on the left side of the body.[FRM+22]

The paper by Fernández et al. [FRM+22] introduces a novel system, ErgoExplorer, designed for interactive visual analysis of ergonomic risk data derived from video streams. ErgoExplorer's importance lies in its ability to process and analyze extensive ergonomic data automatically, making it possible to examine complex relationships between risk assessments of individual body parts over long durations. This capability is crucial in settings like manufacturing and construction, where ergonomic risks are prevalent but challenging to understand, underlining the picture with traditional methods.

The tool utilizes advanced computer vision techniques to extract data from video footage, offering a more systematic and unbiased approach to risk assessment. Its design incorporates multiple views for detailed analysis, including the ErgoView as seen in Figure 3.3 ErgoTimeline and ErgoMovements, facilitating an in-depth understanding of ergonomic risks.

Overall, ErgoExplorer's introduction marks a substantial leap in occupational health and ergonomics, offering a powerful tool for data-driven decision-making in workplace safety. This increased explainability is crucial as it enhances the transparency and comprehensibility of ergonomic assessments, enabling stakeholders to better understand and act on the insights provided, thereby facilitating more effective interventions and promoting safer workplace practices.

The paper by Paudel et al. [PKKC22] proposes a framework which advances ergonomic risk analysis by integrating 3D pose estimation from 2D images or videos. This approach overcomes the limitations of 2D poses, which may not accurately capture complex human movements and postures. The key innovation here is using the 3DMPPE Pose Net method to convert 2D poses to 3D, enhancing the precision of ergonomic risk assessments. A critical component of this framework is the Body Angle Reliability Decision (BARD) network. BARD evaluates the quality of the extracted human pose, determining its suitability for ergonomic scoring. This step is crucial for ensuring that only reliable pose data is used for ergonomic analysis, thereby improving the overall accuracy of the risk assessment.

The shift from 2D to 3D pose estimation offers a more nuanced and accurate understanding of workplace ergonomics. This 3D approach allows for a better understanding of body mechanics and ergonomics, leading to more effective interventions and preventive measures in workplace design and practices.

In conclusion, these advancements in ergonomic analysis technologies represent significant steps in the field, enhancing both the precision and applicability of ergonomic assessments in various workplace settings. Tools like ErgoMaps and ErgoExplorer leverage advanced visualization techniques, transforming complex ergonomic data into intuitive formats like heatmaps and interactive video analysis, facilitating better understanding and decision-making in occupational health and safety. The integration of RULA with motion capture in dentistry and using 3D pose estimation from 2D images in industrial settings are examples of how these technologies are being tailored to specific professional needs. Overall, these innovations signify a move towards more sophisticated, data-driven approaches in ergonomic risk assessment and explainability of the risk, leading to improved workplace design and practices for enhanced worker safety and efficiency.

3.3 Generative approaches & optimization in HPE

This chapter focuses on the application of generative approaches and optimisation in human pose estimation (HPE), a key area in ergonomic research. It particularly examines the use of generative adversarial networks (GANs) and their role in enhancing ergonomic assessment and intervention. The chapter offers a critical review of current literature on utilising GANs in HPE, providing a clear understanding of their capabilities and limitations in ergonomic contexts. Additionally, it explores various optimisation measures crucial for effective ergonomic solutions. Due to the novelty of research in corrective ergonomics and the sparse availability of directly related studies, this chapter includes a selection of papers that, while not identical in focus, contribute to a broader understanding of the field. The aim is to present a comprehensive and coherent overview of these emerging techniques and their potential impact on ergonomic practices.

The paper by El Makrini et al. [EMMV+22] presents a novel approach to enhancing ergonomic safety in human-robot collaborative environments. The methodology uses a virtual spring model to determine the most ergonomic human joint angles, thereby improving worker comfort and reducing the risk of musculoskeletal disorders (MSDs). This is achieved through a framework with a 3D skeleton tracker for real-time posture analysis, a feedback interface for alerting users about non-ergonomic postures, and a workpiece position controller that adjusts the robot's behaviour to promote better human postures. Notably, this approach aligns with the broader theme of corrective ergonomics, where AI optimises human posture. Though the paper primarily focuses on optimising robot pathways for better ergonomics, it aligns closely with the concept of using AI for ergonomic improvements, making it a relevant study in the context of this thesis.

A key aspect of this research is its emphasis on practical implementation in industrial settings. The approach assesses and optimises human postures and integrates seamlessly with the robot's operations, ensuring that ergonomic considerations are a central component of the collaborative process. The methodology was validated through user studies demonstrating the system's effectiveness in improving ergonomic conditions during tasks like object handling, painting, or polishing. Overall, this research is significant for its potential to make human-robot collaboration safer and more efficient. Addressing ergonomic risks in real-time and adapting robotic behaviour accordingly offers a promising solution for reducing workplace injuries and enhancing productivity in various industrial contexts.

The work by Wang et al. [WSH+19]introduces a novel approach in the field of biometric identification, explicitly focusing on gait recognition through a method known as the Two-Stream generative adversarial network (TS-GAN). This method is particularly relevant to research in corrective ergonomics because it demonstrates the potential of GANs to analyze and synthesize human movement patterns effectively. In ergonomics, similar techniques can be applied to assess and enhance ergonomic practices by modelling and optimizing human postures in the workplace. The TS-GAN's ability to distinguish subtle variations in gait under different viewing angles can be adapted to recognize and correct inefficient or unsafe postures from various observational viewpoints. This capability is crucial for developing non-invasive, real-time monitoring systems that improve workplace safety and efficiency, akin to its applications in surveillance and medical diagnostics.

The methodology of the paper centers around addressing the challenge of varying viewing angles in gait recognition. The proposed TS-GAN model innovatively combines global and local feature learning through a dual-stream architecture. This includes a global stream that captures the overall shape and movement and a part stream that focuses on specific body parts, thus enhancing the detail and accuracy of the recognition process. Moreover, the model employs a pixel-wise loss function to improve the fidelity of the generated gait images, ensuring that finer details are not lost during the view transformation process. This combination of global and local feature extraction and high-resolution image generation represents a significant advancement over previous methods, which often struggled with view angle variations and loss of detail.

The efficacy of the TS-GAN model is demonstrated through comprehensive testing on established gait datasets, showcasing its superiority in handling cross-view gait recognition challenges. The results indicate that the TS-GAN not only maintains the integrity of gait features across various views but also outperforms existing state-of-the-art methods in accuracy. This improvement in performance underscores the potential of TS-GAN in real-world applications where view variance is a common issue. The research opens up new avenues in biometric technology, offering a robust and versatile solution for gait recognition that could be pivotal in enhancing both security and medical diagnostics.

In the paper by Xue et al. [XAS+20], researchers present a novel approach that significantly enhances the accuracy and robustness of gait classifiers. The core of this innovation lies in the strategic use of transfer learning and the integration of generative adversarial networks (GANs), specifically through the development of Frame-GAN. The methodology centres around training Frame-GAN on the CASIA-B dataset and transferring the generator's parameters to a Frame-GAN model tailored for the OU-ISIR dataset. This transfer learning technique not only capitalizes on the commonalities in movement modes across different datasets but also remarkably improves the visual quality and effectiveness of the model on the OU-ISIR dataset.

The significance of this study is underscored by its successful implementation of Margin Ratio Loss (MRL), a novel technique designed to boost the classifier's robustness. This method diverges from conventional deep learning strategies that focus on refining the model's structure. Instead, it emphasizes generating higher quality data within the existing dataset, a tactic proven beneficial as evidenced by the enhanced performance over the baseline model, GEINet. Remarkably, the study found that models trained on datasets with artificially increased frame rates (using the Frame-GAN) outperformed those trained on the original frame rate datasets. This outcome illustrates the efficacy of the generated frames in terms of classification accuracy and highlights the potential of Frame-GAN in improving the quality of gait datasets. Innovatively combining transfer learning with GANs and introducing MRL opens a new way for more accurate and robust gait recognition systems. The methodology focuses on enhancing data quality rather than solely relying on deepening network architectures. Furthermore, the promising results from this study encourage future exploration into more complex gait recognition scenarios and the application of this method to larger, more varied datasets.

The paper by Siarohin et al. [SSLS19] primarily focuses on generating human images based on two key variables: the appearance from one image and the pose from another. This task is crucial for various applications, including computer graphics, data augmentation for person re-identification (Re-ID) systems, and human pose estimation technologies. Traditional methods in this domain, such as generative adversarial Networks (GANs) and variational autoencoders (VAEs), have struggled with large spatial deformations between the appearance in the conditioning image and the target pose. This paper addresses these challenges by proposing a novel method that effectively manages these deformations.

The methodology introduced in the paper centres around deformable skip connections within a U-Net-based generator framework. This approach allows for the accommodation of pose-related spatial deformations. The process involves breaking down the overall deformation into a series of local affine transformations, which correspond to subsets of body joints. These transformations are then applied to the convolutional feature maps of the encoder, enabling the accurate transfer of appearance details (such as clothing texture and colour) to the new pose. This method significantly improved over previous approaches, which often failed to handle large spatial displacements effectively, leading to less realistic image generation.

The paper's contribution is noteworthy for its practical implications in augmented reality, virtual try-on applications, and advanced surveillance systems. The ability to generate realistic human images with accurate pose and appearance correlation is vital for enhancing the realism and applicability of virtual models in various domains. Furthermore, by extending the model to include a third conditioning variable - the background - the paper opens up new possibilities for generating more diverse and contextually appropriate images, which is particularly beneficial for creating realistic sequences in video and animation.

This chapter explored various innovative applications of generative adversarial networks in fields adjacent to corrective ergonomics, offering insights that could be adapted to our focus on correcting unergonomic human poses. Though not directly aligned with corrective ergonomics, these studies provide valuable perspectives on using GANs for posture analysis and improvement. Each study presents a unique approach to understanding and optimising human movement, from the use of a virtual spring model for enhancing ergonomic safety in human-robot collaborative environments to the advanced gait recognition techniques for biometric identification and human pose generation. The methodologies discussed, including real-time posture feedback systems, dual-stream architecture for detailed gait analysis, and the use of deformable skip connections for managing spatial deformations in human imaging, demonstrate the potential of GANs in creating more ergonomic and efficient human interactions in various settings. These insights, while stemming from different applications, collectively contribute to the broader narrative of employing AI and GANs for ergonomic advancements, offering a foundational understanding that could be pivotal in developing corrective ergonomic solutions.

CHAPTER 4

Implementation

4.1 From Diagnosis to Action - a Design Rationale

The integration of artificial intelligence into ergonomic assessments, mainly through machine learning techniques such as convolutional neural networks (CNNs) and deep neural networks, represents a transformative step forward in the field of workplace health and safety. These technologies offer precision in the analysis of human posture and motion, leading to ergonomic evaluation tools that pinpoint areas of poor ergonomic practice within the workplace. Despite these advances, a significant gap remains in the transition from diagnostic identification of ergonomic risks to providing actionable, corrective measures. This shortfall highlights the ongoing dependence on human experts for interpreting AI-generated data and implementing corrective actions, indicating the early stages of automation in ergonomic assessments. To address these challenges, this research proposes applying neural style transfer techniques, initially developed for image processing, to the ergonomics field for optimising human posture in workplace settings. By overlaying correct pose sequences onto images capturing unergonomic postures, this approach introduces a novel method for visualising and correcting ergonomic issues beyond mere risk identification to offer practical visual guidance for improvement. This system aims to fulfil the critical need for a comprehensive ergonomic challenge solution, providing identification and correction within a single framework.

Building on the insights from our EPA Loop model [DKS24], which emphasises Evaluation, Problem Identification, and Action, this thesis proposes a structured approach to implementing AI in ergonomic interventions. The EPA Loop model, informed by our position paper on automated ergonomics assessment, underscores the importance of a holistic, user-centric approach beyond traditional diagnostics. It suggests a need for AI solutions that not only detect ergonomic risks but also elucidate them to workers in an understandable manner and guide them towards effective corrective measures. In the context of this research, the Evaluation phase will involve the use of AI techniques to precisely quantify ergonomic risks by analysing posture data collected in situ. This quantitative risk assessment forms the foundation for the subsequent phases of the EPA Loop. During the Problem Identification phase, the research should leverage AI-generated insights to educate workers about the risks associated with their current work postures and practices. Visualisation techniques will be crucial in simplifying complex ergonomic data, enabling workers to identify and comprehend the areas needing improvement easily. This thesis is dedicated to the Action phase, emphasising using neural style transfer for corrective ergonomics. It outlines the creation of interactive guides and personalised corrective strategies through an AI-driven approach. These strategies provide direct, actionable advice for workers to improve their postures and adjust their work habits to mitigate ergonomic risks. This research aims to bridge the gap in ergonomic assessments by integrating the principles of the EPA Loop model with the application of neural style transfer techniques. It seeks to create a novel system that identifies ergonomic risks and offers guidance for improvement.

4.2 Overview

This section outlines the workflow used to apply AI for ergonomic improvements, depicted in Figure 4.4. The workflow starts with taking a video capture of a worker's activities, followed by the extraction of 2D poses. These poses are transformed into 3D using the neural component, after which ergonomic risks are evaluated. The final step involves refining these poses through a multi-component optimisation algorithm to enhance posture while maintaining realism and structural integrity, which presents the main contribution of this thesis.

4.2.1 Retrieval of 3D pose

After taking a video of a person working, an OpenPose framework is used first to extract 2D key points of a human body. Openpose is a robust tool for extracting key points from videos and serves as the foundation for this evaluation process [KSS+21]. The transition from 2D key points to 3D representations is a pivotal step in assessing ergonomic risks with heightened accuracy and depth. The 2D to 3D lifting is achieved by leveraging a deep learning model, Videopose3D [PFGA19], trained to infer the third dimension from the spatial configurations and dependencies observed in the 2D key points. The model utilises a convolutional neural network architecture predominantly comprising convolutional layers, specifically enhanced with dilated convolutions, to predict the 3D coordinates of each key point from 2D video frames. This architecture is adept at handling the temporal sequences and spatial hierarchies inherent in human movement. It explicitly incorporates anatomical constraints and relative positioning of body parts, leveraging the expansive temporal receptive field afforded by dilated convolutions to capture long-term dependencies without needing fully connected layers. The result is a sequence of 3D key points for each frame, accurately reflecting the three-dimensional posture of the person being analysed, ensuring that the predicted poses adhere to realistic human

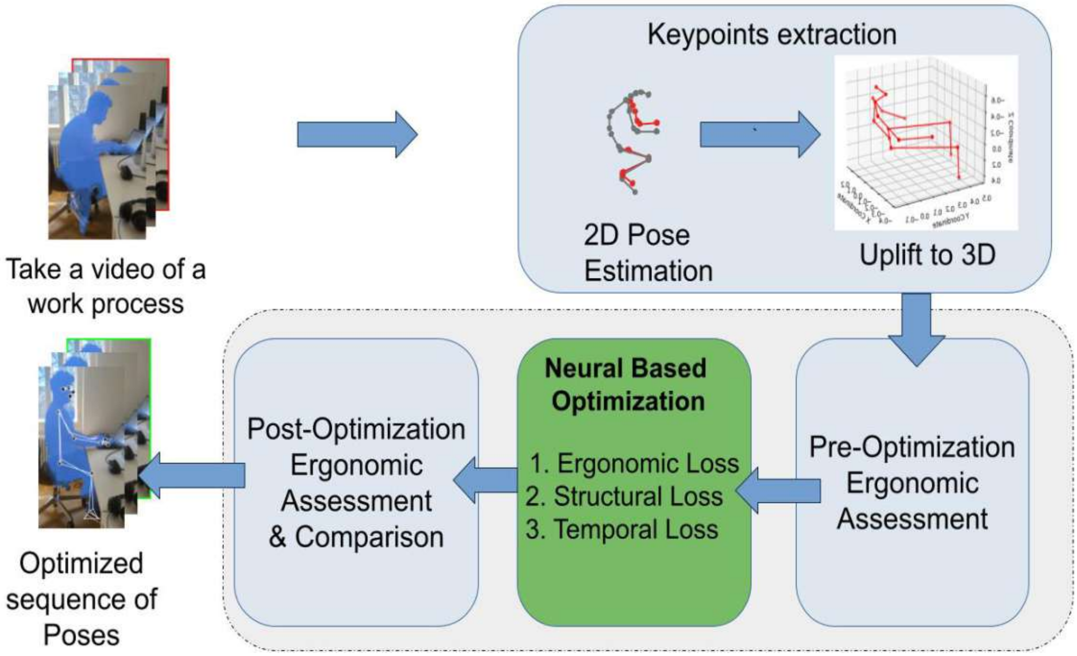


Figure 4.1: Proposed process of ergonomic correction

biomechanics. This 3D pose information is required for the ergonomic assessment by providing a comprehensive view of the worker's posture in the workspace. It allows for a more nuanced analysis of ergonomic risk factors, such as awkward postures or movements that may not be fully apparent in 2D projections. Moreover, the 3D representation facilitates the calculation of more accurate ergonomic risk scores, such as the RULA (Rapid Upper Limb Assessment) score, by considering the spatial orientation of the body parts. Consequently, the transition from 2D to 3D key points is a critical enhancement in the methodology of automated ergonomic assessment.

4.3 Optimisation Method

This section focuses on optimising human postures to ensure they are ergonomically practical and structurally sound while maintaining a realistic appearance. The goal is to minimise the overall loss associated with ergonomic deviations and structural inconsistencies, enhancing the realism of human poses through a systematic optimisation algorithm. The general mathematical formulation of the optimisation process can be described as follows:

$$\text{Composite Loss} = \min\left[\alpha \cdot L_{\text{Ergo}}(\text{kps}) + \beta \cdot L_{\text{Struc}}(\text{kps}) + \gamma \cdot L_{\text{Disc}}(\text{kps})\right] \quad (4.1)$$

Where:

- **kps** are 17 keypoints with 3 dimensions x,y,z

- **Ergonomic Loss**: The ergonomic loss, denoted as $L_{\text{Ergo}}(\text{kps})$, assesses the deviation of the current posture from an ideal ergonomic position. This function is typically modelled using RULA or similar ergonomic assessment methods. It is designed to guide the optimisation process toward postures that minimise the risk of musculoskeletal issues. By quantitatively assessing how well a pose aligns with ergonomic principles, this loss helps identify postures that could potentially lead to discomfort or injury over time. The goal is to adjust the pose to minimise ergonomic loss, indicating an alignment with ergonomic best practices.

- **Structural Loss**: The structural loss, $\mathcal{L}_{\text{struc}}(\text{kps})$, ensures that the key points between frames remain in a physically plausible arrangement, maintaining the integrity of the human pose over time. It focuses on the temporal consistency of the pose sequence, aiming to reduce flickering between frames and ensure smooth transitions.

- **Discriminator Loss**: The discriminator loss, $L_{\text{Disc}}(\text{kps}$, penalizes non-human-like poses. This component, derived from a generative adversarial network, helps ensure that the optimised poses are realistically human-like. It functions as a quality check against unrealistic poses, with the discriminator model identifying and penalising poses that do not adhere to human biomechanical constraints. The discriminator's feedback helps refine the optimisation process, ensuring that the resulting poses meet ergonomic standards and are visually convincing.

- **Alpha (α)**: The weight associated with the ergonomic loss. It determines the emphasis placed on minimising the deviation of the current posture from an ergonomically ideal posture.

- **Beta (β)**: The weight assigned to the structural loss. It controls the importance of maintaining the integrity of the human pose between frames, ensuring smooth transitions and temporal consistency.

- **Gamma (γ)**: The weight linked to the discriminator loss. This parameter adjusts the focus on ensuring that the poses are realistically human-like, as determined by a GAN-based discriminator.

The goal is to find the key configuration that minimises the overall loss, combining ergonomic risk, structural feasibility, and human likeness. The optimisation begins with an initial pose represented in 3D space, which is then cloned and set to require gradients, allowing for applying optimisation algorithms. The optimiser, chosen from various options such as Adam [KB14], is responsible for adjusting the pose parameters based on the calculated gradients to minimise the loss function, where α, β, and γ are the weights for the ergonomic, structural, and discriminator losses, respectively. These weights are adjustable parameters that balance the contribution of each loss component to the overall optimisation process. Each loss component plays a crucial role in the overall optimisation framework, contributing to the development of an automated system capable of generating ergonomically optimised and realistic human poses. A higher α tilts the optimisation towards enhancing ergonomics, potentially at the expense of pose realism or smoothness. Conversely, an elevated β value emphasises the fluidity of motion, which might inadvertently sideline ergonomic optimisation or realism. Lastly γ, a dominant prioritises the human-like appearance of poses, which could detract from ergonomic adjustments or smooth transitions. Optimal results are achieved by meticulously adjusting these hyperparameters, ensuring that no single aspect is improved in isolation but rather in harmony with others. In the process of optimising poses for ergonomic improvements, adjusting the weights (α, β, and γ) associated with each type of loss (ergonomic, structural, and discriminator loss, respectively) serve as levers to control the emphasis placed on reducing ergonomic risks, maintaining structural integrity, and ensuring the human-likeness of the optimised poses. Throughout this research, a variety of weight configurations have been explored to understand their impact on the optimisation results. This iterative testing is driven by the hypothesis that different tasks or movements might require distinct balances between ergonomic safety, anatomical accuracy, and realistic movement patterns. For instance, a task involving repetitive motion might prioritise ergonomic loss reduction more heavily, necessitating a higher weight for α compared to tasks that are less repetitive but require precise control, where structural integrity β might be more critical. This exploration is guided by quantitative metrics, such as the reduction in ergonomic risk scores and qualitative assessments of pose realism and feasibility. The ultimate goal is to establish a set of guidelines or adaptive algorithms that can intelligently adjust the weights based on the specific requirements of the task at hand, the initial ergonomic assessment, and the progress of the optimisation, thus paving the way for more personalised and effective ergonomic interventions.

By reducing the influence of ergonomic loss on the overall loss function, the network was encouraged to make more conservative adjustments to poses, thereby preserving more of the original movement intent. This approach provided a temporary solution, allowing for improved ergonomic scores without as drastic a change in the movement's nature as before. However, this solution has its shortcomings. While it reduces the likelihood of significantly altering the intent of the movement, it also limits the extent to which ergonomic improvements can be achieved. The balance between maintaining movement intent and optimising for ergonomics becomes a delicate one, with the reduced weight on ergonomic loss possibly compromising the effectiveness of ergonomic optimisation.

The optimisation process is applied to a series of poses extracted from video data, demonstrating the method's scalability and potential for widespread application in ergonomic assessments and corrections. By leveraging the neural style transfer technique in a novel context, this thesis proposes a new approach to automating and enhancing ergonomic assessments, merging the fields of artificial intelligence, ergonomics, and computer vision.

4.3.1 Ergonomic Loss

The first step in the ergonomic optmisation process involves evaluating the current pose to ensure it is conducive to a gradient-based optimisation. This is achieved by tensorizing all calculations of the RULA method, which traditionally uses non-differentiable step functions. To facilitate continuous optimisation, the approach here utilises differentiable sigmoid functions to mimic these traditional step functions. Rapid Upper Limb Assessment (RULA) is a well-established method used to evaluate the risk of musculoskeletal disorders resulting from repetitive movements and static postures of the upper body [MC93]. However, the traditional RULA approach is not directly applicable to neural-based optimisation processes due to its reliance on step functions for score determination (see 4.2). This characteristic poses a significant challenge when integrating RULA within an automated, AI-driven framework, as neural networks require differentiable functions to optimise through gradient descent.

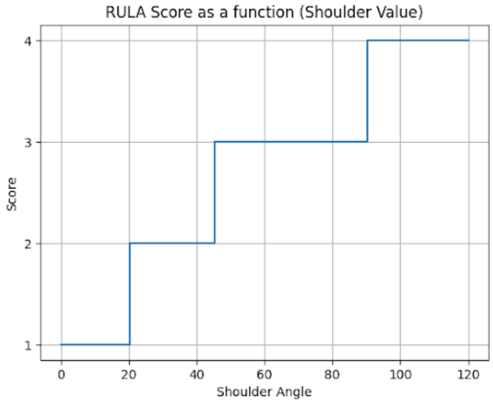


Figure 4.2: Representation of RULA step functions

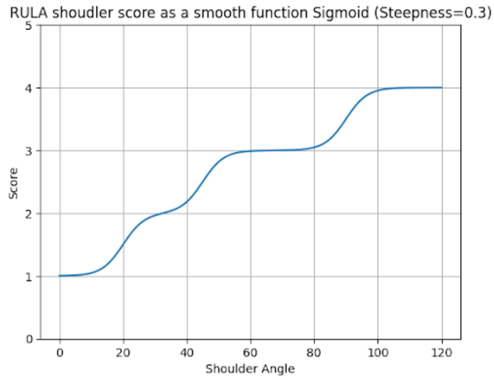


Figure 4.3: Diferantable representation of RULA

To address this limitation, this thesis has introduced an approach using sigmoid functions as surrogates for the step functions inherent in RULA scoring. Sigmoid functions are smooth, differentiable, and can be tuned to approximate the thresholds and transitions between RULA scores, thus making them compatible with gradient-based optimisation techniques (see 4.3). This method allows for continuous, real-time assessment of ergonomic risks by providing a differentiable pathway to adjust and optimise body postures towards ergonomically favourable positions.

The sigmoid-based scoring system is designed to reflect the sensitivity of different body parts to postural changes. Parameters such as the steepness and position of the sigmoid transitions are carefully calibrated to mirror the scoring dynamics of the RULA method. This adaptation ensures that the neural network can effectively learn and propose adjustments to the detected poses, aiming to minimise ergonomic risks without losing the integrity and reliability of traditional RULA assessments. The function is defined as:

$$S(x) = \frac{1}{1 + e^{-k(x-x_0)}} \tag{4.2}$$

Where:

- $S(x)$ is the output of the sigmoid function,
- x is a posture angles calculated from the extracted 3D pose
- k is the steepness of the sigmoid curve, determining how rapidly the function transitions from 0 to 1 around the point x_0,
- x_0 is the position along the x-axis where the transition (i.e., the midpoint of the sigmoid) occurs, representing the threshold between different RULA scores.

In the context of the ergonomic scoring system, the parameters k and x_0 are of critical importance:

- **Steepness (k)**: This parameter controls the sensitivity of the scoring function to changes in posture. A higher value of k results in a steeper transition, meaning the scoring system quickly shifts from one score to another as the input x crosses the threshold x_0. This steepness is crucial when adjustments to posture need to be highly responsive to minor deviations from ergonomic norms.
- **Position (x_0)**: This parameter defines the critical point at which the score transitions. By adjusting x_0, the model can set appropriate thresholds that correspond to the original step functions in RULA, where each step represents a different level of ergonomic risk.

The integration of sigmoid functions into the RULA evaluation framework enables the continuous optimisation of body postures in a tensorised way.

4.3.2 Structural Loss

The structural loss, $\mathcal{L}_{\text{struc}}(\text{kps})$, ensures that key points between frames remain in a physically plausible arrangement, thereby maintaining the integrity of the human pose over time. It focuses on temporal consistency in pose sequences, aiming to reduce flickering between frames and ensure smooth transitions. This loss is significant because the optimisation treats each frame's pose individually without knowledge of the next frame, leading to potential flickering. The loss, calculated using the Temporal Mean Squared Error (TMSE) [Hur88], evaluates squared differences between adjacent frames to promote consistency and prevent jarring visual effects. The TMSE formula is:

$$TMSE(\mathbf{p}) = \frac{1}{N-1} \sum_{i=1}^{N-1} (\mathbf{p}_{i+1} - \mathbf{p}_i)^2 \tag{4.3}$$

Where p_i and p_{i+1} are positions in the array optimised poses at indices i and $i+1$ respectively, and N is the total number of frames in the sequence.

4.3.3 Discrimnator Loss

This section discusses the implementation of generative adversarial network architectures, focusing on the use of discriminator networks in the optimisation process. The main goal was to develop a generative adversarial network framework that effectively utilises models to generate and discriminate against human poses. Figure 4.4 depicts the general architecture of Human Pose GAN.

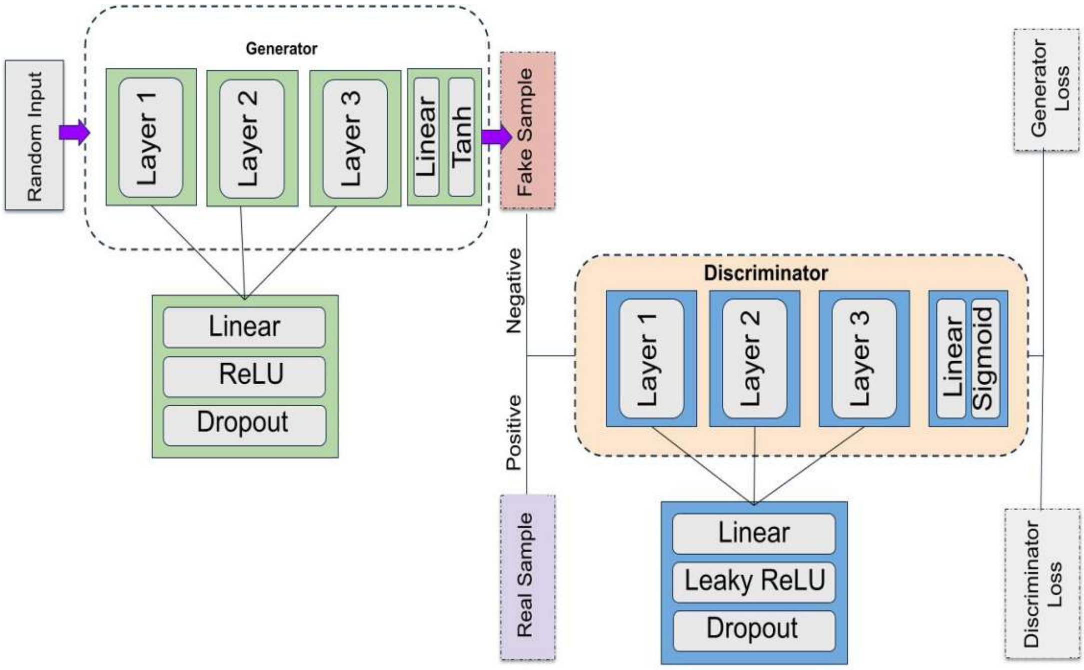


Figure 4.4: Scheme of the proposed Human Pose GAN

Human Pose Discriminator

The discriminator's role is to distinguish between real and generated poses, providing feedback to the generator on the realism of its outputs. The initial discriminator design consisted of a sequence of linear layers, LeakyReLU activation functions for non-linearity, and dropout layers to prevent overfitting. This setup was crafted to evaluate the detailed features of pose representations effectively. To further enhance the discriminator's ability to assess poses, experiments with adding more layers and adjusting dropout rates were conducted. The introduction of L1 regularisation was a strategic choice aimed at encouraging the model to focus on the most salient features of the poses, potentially improving its discrimination capability. Moreover, enriching the discriminator's input with angle and structural information derived from the poses was a novel approach to incorporating more ergonomic context into the evaluation, simulating how human experts assess posture more closely. This model acts as the discriminator in the GAN setup, evaluating whether the input poses are realistic (i.e., resembling true human poses). Its architecture is composed of:

- **Input Layer:** Accepts a 51-value input corresponding to 17 keypoints, each with x, y, and z coordinates.
- **Hidden Layers:** Consists of several fully connected (dense) layers with decreasing units from 128 to 1. Each hidden layer employs LeakyReLU for activation, promoting non-linearity and helping to avoid vanishing gradient issues during training. Dropout is included after each activation to prevent overfitting by randomly setting a portion of the feature detectors to zero during forward and backward passes.
- **Output Layer:** A single unit with a sigmoid activation function produces a probability between 0 and 1, indicating the likelihood that the input pose is realistic.

Human Pose Generator

The generator's primary function is to synthesize human poses that are realistic and ergonomically possible. The network architectures explored range from simpler, fully connected models to more complex designs incorporating convolutional layers. Initially, a sequence of linear layers with ReLU activation functions and dropout for regularization was utilized. This setup aimed to map a latent space vector to a pose representation. However, to enhance the generator's capability to produce more detailed and varied poses, the architecture was expanded to include additional layers and experimented with different sizes and arrangements. These variations were tested to find an optimal balance that could generate detailed pose representations without causing overfitting or making the training process unstable. For example, expanding the network with more layers aimed to increase the model's capacity to learn complex pose features but also introduced challenges in training stability and convergence. This model serves as the generator in the GAN setup, tasked with creating human poses from random noise input. The final design of generator:

- **Input Layer:** Starts with a 100-dimensional noise vector.
- **Hidden Layers:** Similar to the discriminator, it uses fully connected layers but in an increasing manner from 128 to 512 units. Each layer utilizes ReLU activation for non-linearity and includes dropout where specified to aid in generalization.
- **Output Layer:** Ends with a 51-unit layer using the Tanh activation function to generate the 3D coordinates for each of the 17 keypoints. Tanh is used to ensure

the output values are normalized between -1 and 1, which typically matches the normalized range of human pose coordinates.

The discriminator uses binary cross-entropy loss for its real/fake classification task, and the generator's effectiveness is measured by how well it can trick the discriminator.

Additionally, experiments with convolutional generative adversarial network architectures were perfomed. This represented an attempt to leverage the spatial processing capabilities of convolutional neural networks. Convolutional discriminator and generator models were designed to process pose data in a manner that respects the spatial relationships between keypoints. These models employed convolutional layers to extract features from the pose data, aiming to capture the complex spatial hierarchies inherent in human poses.

However, despite the theoretical advantages of convolutional architectures in processing spatial data, the experiments with convolutional generative adversarial network architectures did not yield the expected improvements. This outcome suggests that the unique characteristics of pose data and ergonomic assessment criteria may require more specialized network designs or feature representations than those captured by standard convolutional layers.

The iterative exploration of network architectures for both the generator and discriminator highlighted the complexities involved in designing effective GANs for ergonomic pose optimization. Each variation in the network configuration—whether it involved adjusting layer depths, exploring different activation functions, or incorporating novel input features—was driven by the goal of achieving a delicate balance. This balance had to ensure the generation of realistic, varied, and ergonomically improved human poses while maintaining model stability and avoiding overfitting. Through extensive experimentation, including trials with convolutional architectures, valuable insights were gained into the nuanced requirements of neural network design for specific applications like ergonomic pose optimization. These insights not only inform the ongoing development of GANs for this purpose but also contribute to the broader field of AI-driven ergonomics assessment.

4.3.4 Results of the Human Pose GAN

In the development of generative adversarial networks (GANs) for the creation of human poses, significant advancements have been observed, particularly in the realism and physical feasibility of the movements generated. This progression has been prominently highlighted through the process of comparing the GAN outputs with real pose data, complemented by the visualization of these poses across different training epochs. Such visual assessments have demonstrated a steady improvement in the generator's outputs, evolving to produce increasingly life-like poses. Simultaneously, enhancements in the discriminator's ability to discern between real and synthetic images underscore the success of the adversarial training strategy. A key indicator of this success is the discriminator's prediction rates approaching a balanced score of 0.5, suggesting an optimal performance where the discriminator equally identifies images as real or generated, thus validating the model's effectiveness in generating convincingly realistic human poses. This journey towards achieving a delicate balance in the discriminator's assessments marks a pivotal achievement in the project, emphasizing the capability of GANs to mimic intricate human movements accurately.

Through training and parameter optimization, the GAN model has been assessed based on its ability to produce images that mimic real human poses with high fidelity. Visualization of both generated and real images over various epochs provided a view of the model's performance, revealing significant advancements in the realism of the poses and the discriminator's ability to accurately classify the images.

The initial outputs of the GAN, as referenced in Figure 4.6, showed that the generated poses were quite random, indicating the early stage of learning where the generator lacked the capability to produce lifelike images. Over time, with ongoing training as shown in subsequent figures (refer from Figure 4.7 to Figure 4.10), the generated images gradually improved, reflecting a more realistic representation of human poses. This progression highlights the learning curve of the generator in capturing the nuances of human movements. Simultaneously, it is evident from Figure 4.9 that although the GAN at epoch 100 generated a pose that was not quite human-like, the discriminator accurately predicted this, reflecting its increased proficiency in identifying such discrepancies. This means, the discriminator evolved to effectively distinguish between non-human and actual human poses, which is a testament to the adversarial training method employed. The discriminator's learning trajectory, improving its predictions to near perfect balance around the 0.5 probability mark by later epochs, underscores its critical role in the GAN framework for ensuring the generation of plausible human-like images.

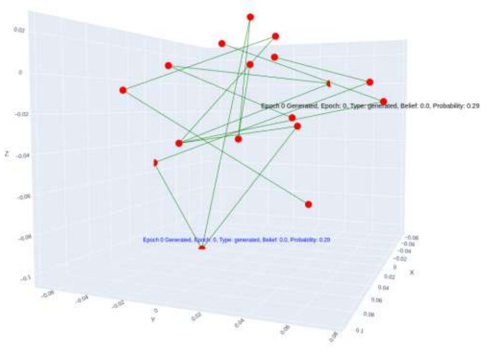
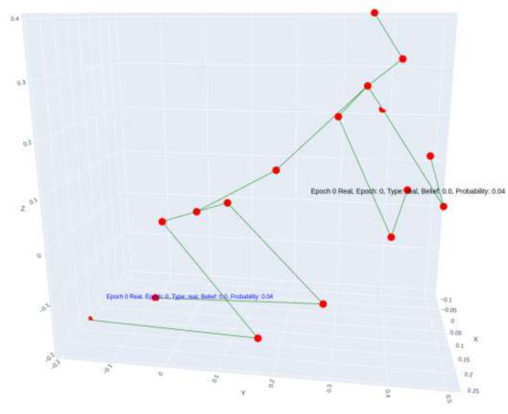
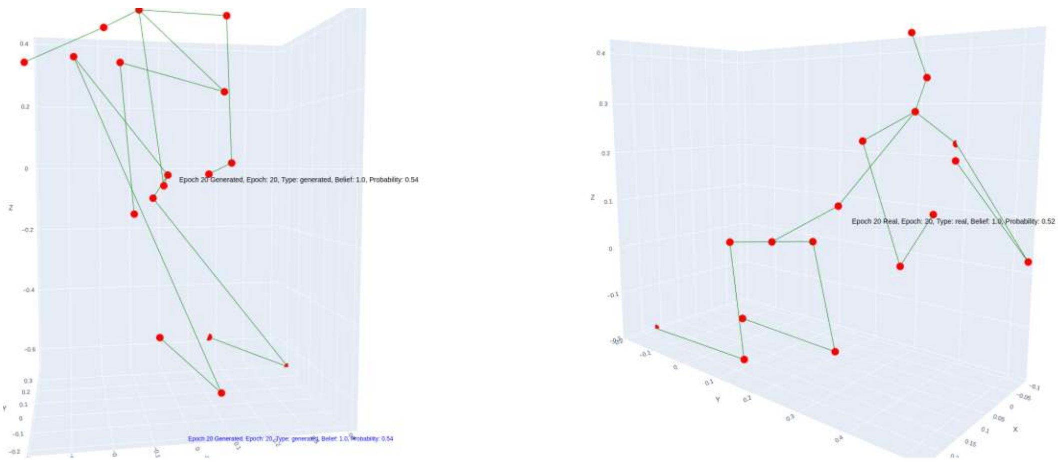


Figure 4.5: Generated image. Discriminator predicts as generated with a probability of 0.29.



Real image. Discriminator predicts as generated with a probability of 0.04.

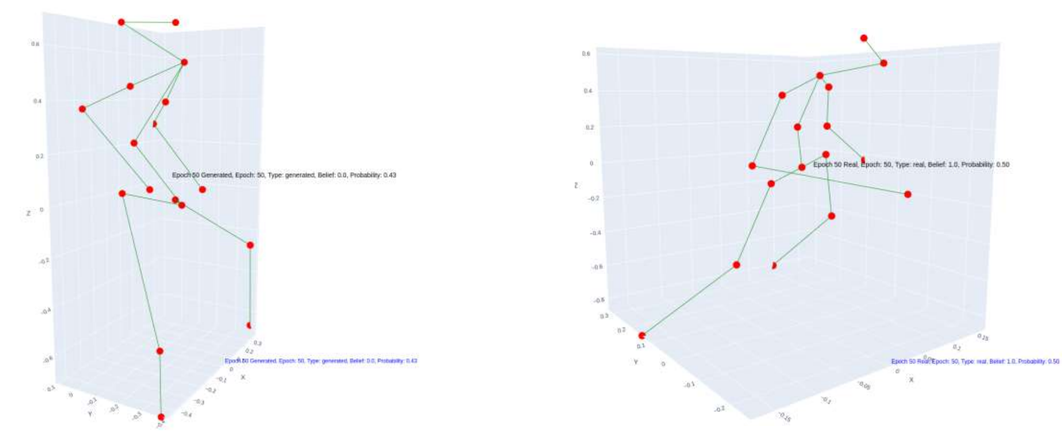
Figure 4.6: Epoch 0: Discriminator responses.



Generated image. Discriminator predicts as real with a probability of 0.54.

Real image. Discriminator predicts as real with a probability of 0.52.

Figure 4.7: Epoch 20: Discriminator responses.

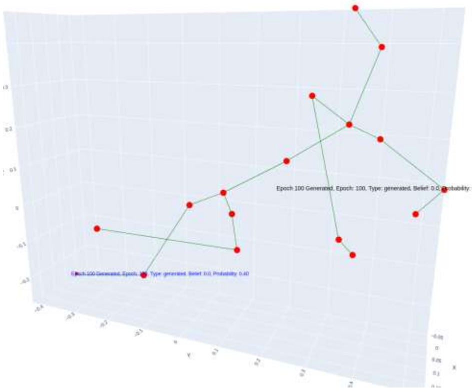


Generated image. Discriminator predicts as generated with a probability of 0.43.

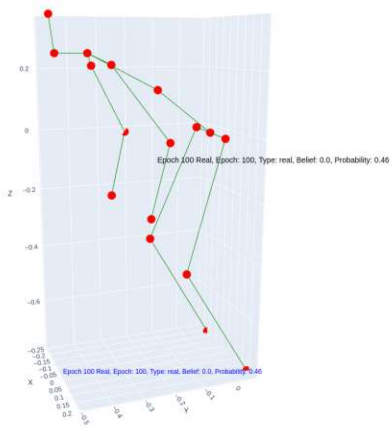
Real image. Discriminator predicts as undecided with a probability of 0.50.

Figure 4.8: Epoch 50: Discriminator responses.

In conclusion, the training and evaluation processes have collectively validated the GAN model's effectiveness in generating realistic and diverse human poses. The approach's success is marked by the achievement of a balanced state of discriminator predictions, reflecting the GAN's ability to generate convincing images that closely mirror real human poses. This balance between generation and discrimination, achieved through careful parameter optimization and rigorous training, demonstrates the GAN model's proficiency in capturing the nuances of human movement. This thesis explores the integration of

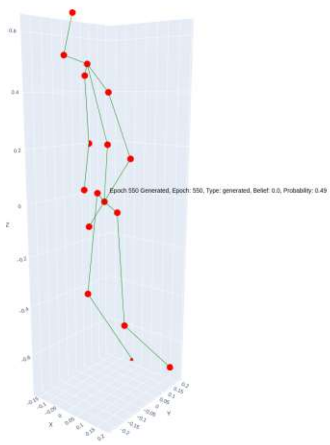


Generated image. Discriminator predicts as generated with a probability of 0.40.

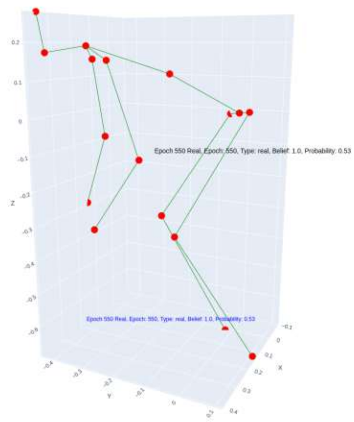


Real image. Discriminator predicts as generated with a probability of 0.46.

Figure 4.9: Epoch 100: Discriminator responses.



Generated image. Discriminator predicts as generated with a probability of 0.49.



Real image. Discriminator predicts as real with a probability of 0.53.

Figure 4.10: Epoch 550: Discriminator responses.

neural style transfer and generative adversarial networks (GANs) for optimizing ergonomic postures in workplace environments. A key challenge in this domain is developing AI models that accurately simulate and enhance human poses for ergonomic safety without falling prey to issues like overfitting or training instability. The discriminator component of GANs plays a pivotal role in this context. It critically evaluates the realism of generated poses, ensuring they closely resemble human-like forms. This capability is vital because maintaining human resemblance in optimized poses enhances the model's practical application.

Moreover, this research underscores the practical implications of integrating traditional ergonomic assessment tools, such as the RULA score, into a format suitable for gradient-based optimization. This transformation introduces complexity, but it is a necessary step to ensure accurate and computationally feasible ergonomic evaluations. The use of GANs, particularly the discriminator's loss function, is instrumental in refining the model to produce ergonomically sound and visually credible postures, thereby facilitating safer workplace environments.

Sigmoid functions were introduced as surrogates for the step functions in RULA scoring, facilitating the integration of ergonomic assessments into an AI-driven optimization framework. This adaptation allowed for real-time, continuous adjustment of postures towards ergonomically favorable positions. In the development of GAN architectures, extensive experimentation with network configurations led to insights on achieving a balance between model complexity and practical efficacy. Adjustments in the depth and composition of the neural networks, alongside the introduction of regularization techniques and enriched input features, helped improve the realism and ergonomic quality of generated poses.

CHAPTER 5

Results and Evaluation

This chapter goes into a detailed evaluation and results of the proposed method, highlighting the impact of various hyperparameters on the optimization process and the significance of each component of the loss function. The method's efficacy is demonstrated through both quantitative metrics and qualitative assessments, including pre-and post-optimization comparisons of ergonomic scores, joint angles, and visual realism. The core of the optimization framework is built upon a three-part loss function: ergonomic loss, structural loss, and discriminator loss. The ergonomic loss is quantified using the RULA (Rapid Upper Limb Assessment) score, a widely recognized ergonomic evaluation method that assesses the risk of musculoskeletal disorders. The structural loss, aimed at reducing flickering between frames, is calculated using the Temporal Mean Squared Error (TMSE) to ensure smooth transitions. Lastly, the discriminator loss, derived from a human pose discriminator, is utilized to maintain the human likeness of the optimized poses. Hyperparameters α, β, and γ from Formula 4.1 balance these components, influencing the optimization's focus on ergonomics, smoothness, and realism, respectively. This interplay of α, β, and γ is central to the framework's effectiveness.

The chosen coefficients for the loss function components α (0.3) for ergonomic loss, β (0.4) for structural loss, and γ (0.3) for discriminator loss were empirically determined to yield the most effective balance between ergonomics, structural integrity, and realism. This configuration reflects a targeted approach to mitigate specific challenges associated with pose generation, such as ensuring ergonomic viability without compromising the natural appearance of the poses.

This section presents the results and evaluation from analysing 30 videos of manual handling tasks. Although 30 videos were processed to gather comprehensive data, the results here pertain specifically to one video that illustrates the critical action of picking up and carrying items due to its significant ergonomic risk factors, particularly regarding back safety. The efficacy of pose optimisation in this scenario is demonstrated through a series of graphs, numbered from Figures 5.1 and 5.5, which document changes in joint angles and ergonomic scores, accompanied by before-and-after visual comparisons of optimised poses. This focused approach highlights the potential improvements in ergonomics that can be achieved through targeted intervention in high-risk activities.

Visually, the comparison between initial and optimized poses was facilitated through the generation of images depicting the poses before and after optimization, providing a clear visual representation of the improvements in pose ergonomics and realism as seen in Figure 5.5. Quantitatively, the thesis relied on the RULA score, a widely recognized ergonomic assessment tool, to evaluate the ergonomic improvement of poses. Additionally, the analysis included measurements of joint angles (e.g., trunk, upper arm, knee angles) to assess the physiological impact of the optimization on pose ergonomics.

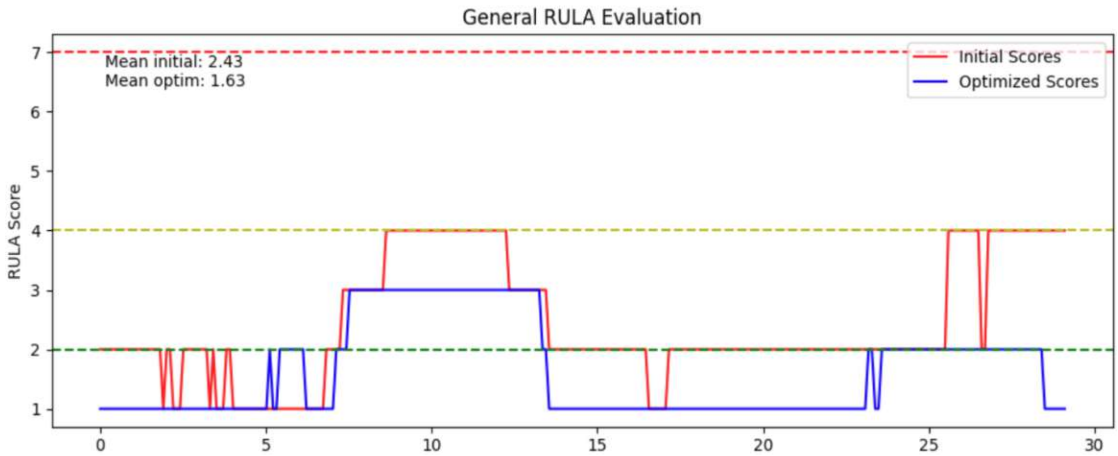


Figure 5.1: General RULA Ergonomic Score Pre and Post-Optimization: This graph illustrates the optimization process in enhancing ergonomic safety. A clear reduction in RULA scores post-optimization signifies a decrease in ergonomic risk.

To further investigate the underlying factors contributing to the overall decrease in ergonomic risk, an in-depth analysis of shoulder angle adjustments was conducted. The examination of shoudler angles, specifically Figure 5.2, revealed the most substantial reduction in shoulder angles among the studied body parts. This significant decrease in shoulder angles is hypothesized to be a primary contributor to the observed reduction in the general RULA score, as shoulder positioning plays a crucial role in the ergonomic assessment of upper body postures.

In addition to the improvements in shoulder angles, further analyses were conducted on trunk and knee angles, which are integral components of an ergonomic assessment. The thesis's findings, illustrated in Figure 5.3, show a consistent decrease in trunk angles post-optimization, suggesting an enhancement in the posture of the lumbar region towards a less strenuous configuration. This adjustment in trunk posture directly contributes to a lower ergonomic risk by minimizing the potential for back-related injuries, which are prevalent in workplace settings.

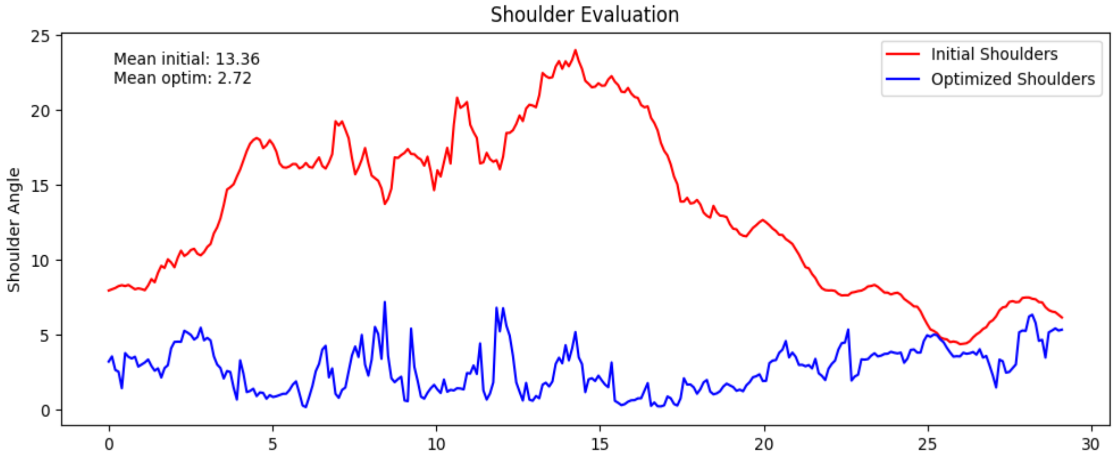


Figure 5.2: Detailed Analysis of Shoulder Angle Adjustments Post-Optimization: significant reduction in shoulder angles, indicating an essential factor in the overall decrease in ergonomic risk.

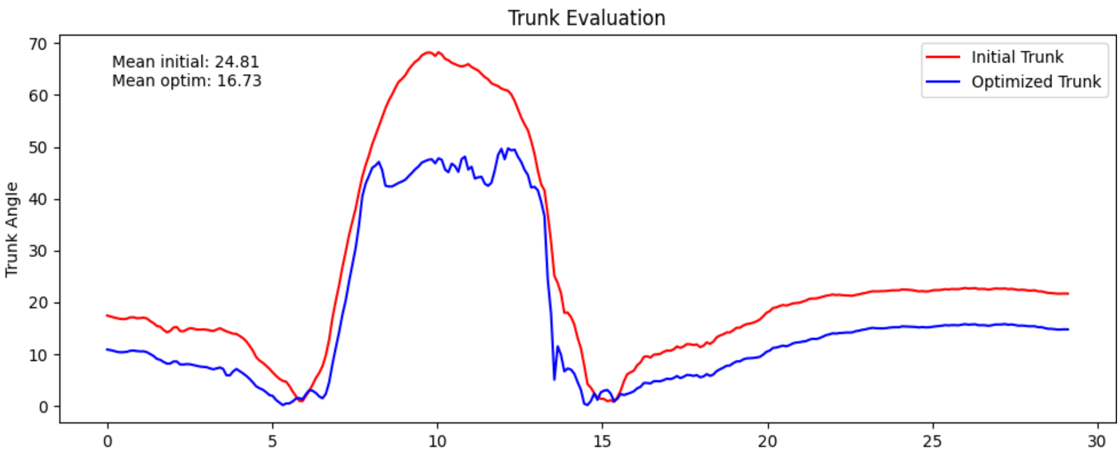


Figure 5.3: Optimization Impact on Trunk Angles: The graph demonstrates a consistent reduction in trunk angles, indicating improved lumbar posture and reduced back injury risk.

The examination of knee angles, as captured in Figure 5.4, revealed both a general trend of increased angles and the presence of sporadic spikes. This variability can be attributed to the interconnected nature of the human body, where adjusting one segment can inadvertently impact others. Specifically, increasing knee angles may facilitate a compensatory reduction in trunk angles, thereby adopting a more ergonomically favorable posture that distributes strain more evenly across the lower and upper body. To visually demonstrate the impact of the optimization process on pose adjustments, a comparison visualization was created as seen in Figure 5.5, pre-optimized poses with their post-optimization counterparts. In this visualization, the post-optimized pose is depicted

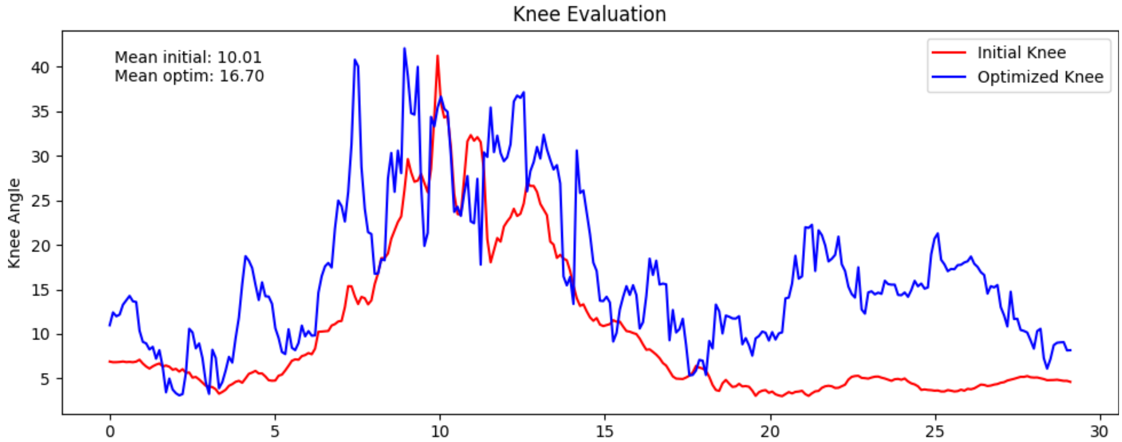


Figure 5.4: Knee Angle Adjustments and Observations: This figure showcases the increase in knee angles with occasional spikes, highlighting the dynamic interplay between different body segments in achieving ergonomic posture.

on the left in green, while the pre-optimized frame is shown on the right. Although the differences between the two poses may initially appear subtle to the naked eye, the incorporation of a black line overlay provides a clearer indication of the changes, particularly highlighting a reduced trunk angle in the optimized pose. This visual aid not only reinforces the quantitative findings discussed earlier but also offers an intuitive understanding of how minor adjustments can significantly enhance ergonomic posture, emphasizing the method's capability to refine worker positions for better health outcomes. Quantitative analyses reveal a notable improvement in ergonomic scores post-optimization. The mean RULA score, a critical indicator of the ergonomic risk associated with a pose, demonstrated a significant decrease, indicating a reduction of ergonomics scoress as seen in Figure 5.2. Furthermore, the quantitative improvements observed, the mean ergonomic score experienced a substantial decrease from 3.2 to 1.8 post-optimization as seen in Table 5.1. This decline signifies not only enhancement in the ergonomic quality of the poses but also a significant reduction in the potential for musculoskeletal strain and injury. The transition towards more ergonomically favorable poses illustrates the effectiveness of the optimization methodology in mitigating ergonomic risks in the workplace.

Ergonomic Assessment	**Pre-Optimization**	**Post-Optimization**
Mean Ergonomic Score	3.2	1.8

Table 5.1: Comparison of Mean Ergonomic Scores Before and After Optimization

The results demonstrated notable improvements in ergonomic scores and joint angles, indicating a successful optimization of poses towards ergonomically favorable configurations. Specifically, the optimization led to an increase in knee angles and a reduction in trunk and upper arm angles, aligning with ergonomic guidelines for minimizing strain and improving posture. However, the thesis acknowledges that while the quantitative im-

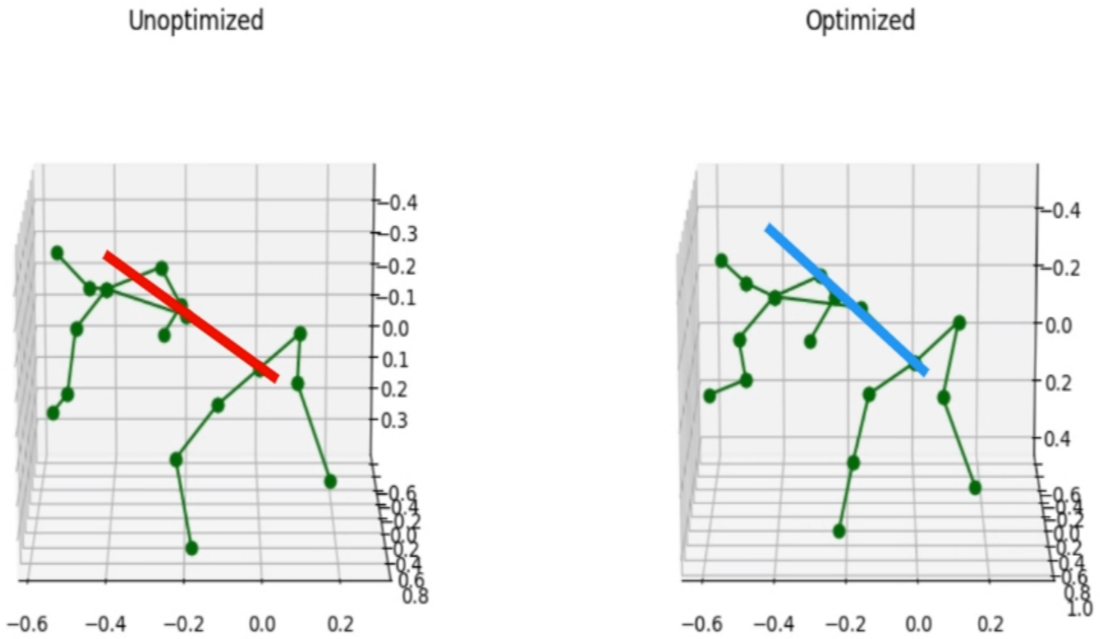


Figure 5.5: Pose Comparison Visualization: This image compares pre-optimized and post-optimized poses, with a black line overlay emphasizing the reduction in trunk angle post-optimization.

provements are significant, the subtlety of changes might not be immediately perceptible to the human eye.

Despite the overall success of the optimization process in enhancing ergonomic postures, it is crucial to acknowledge instances where the outcome may not align with expectations. In certain cases, the optimization could inadvertently lead to less desirable poses, underscoring the complexity of human biomechanics and the challenges inherent in automated optimization processes. Such outcomes highlight the importance of more research and refinement to ensure the efficacy and safety of ergonomic interventions.

Following the presentation of examples where the optimization process did not yield the expected ergonomic improvements, as seen in Figure 5.6, it is essential to classify these instances into four distinct cases to better understand the limitations and challenges of the current methodology:

1. **Insignificant Ergonomic Difference:** In some cases, the optimization process yields minimal or indiscernible improvements in ergonomic scores. The poses before and after optimization exhibit no significant change, rendering the optimization ineffective from an ergonomic perspective.

2. **Structurally Incoherent Poses:** The optimization algorithm sometimes generates

poses that appear unnatural or messy deviating from human-like structures. This can result from the model prioritizing certain ergonomic adjustments without maintaining the overall coherence of the human pose.

3. **Disproportionate Adjustment of Unaccounted Points:** Optimization focused on enhancing specific key points for ergonomic scores can inadvertently shift other points of the human pose that are not directly considered in the calculation. This can lead to unnatural postures and misalignments not initially intended.

4. **Loss of Initial Movement Intent:** In efforts to optimize for ergonomics, the original intent of the movement (e.g., lifting) can sometimes be obscured or lost. The optimized pose might prioritize ergonomic efficiency at the expense of the functional aspect of the movement, leading to a compromise in the activity's effectiveness or purpose.

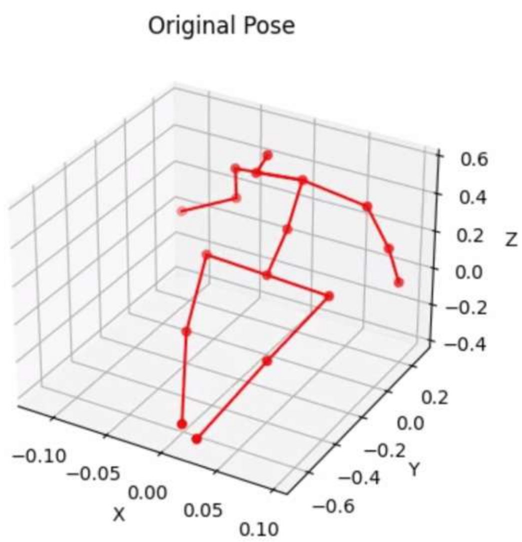


Figure 5.6: Example of Suboptimal Optimization Outcome: This comparison illustrates a sitting scenario where the optimization process resulted in a pose that did not achieve the intended ergonomic improvement.

Identifying and categorizing these instances provides a clear direction for refining the optimization process. Addressing these challenges will enhance the model's ability to produce both ergonomically optimized and realistically human-like poses, ensuring the practical applicability of the methodology in real-world scenarios.

This research contributes to the field of human pose generation by demonstrating the feasibility of incorporating ergonomic optimization into the GAN framework. The successful integration of ergonomic considerations into the pose generation process not only enhances the realism and safety of the generated poses but also underscores the potential of machine learning techniques in creating ergonomically optimized digital human models.

CHAPTER 6

Discussion

This chapter delves into the achievements and challenges encountered in applying machine learning techniques, specifically generative adversarial networks and neural style transfer, to ergonomic assessments. It highlights significant advances in enhancing the efficiency and accuracy of ergonomic posture analysis and outlines innovative solutions proposed to overcome existing limitations. Additionally, this section proposes future directions for research aimed at refining and expanding upon the preliminary successes documented in this thesis. Through this discussion, this thesis aims to present a balanced view of the current state of research, acknowledging its strengths and addressing its shortcomings while paving the way for future advancements in the field. The integration of neural style transfer with gradient-based optimization has enabled the generation of human-like poses that adhere to ergonomic principles and has also provided insights into the optimization processes underpinning neural networks. A profound understanding of ergonomic assessments and their critical role in maintaining workplace health and safety was developed throughout this research. The thesis illuminated the complex nature of ergonomic risk factors and the necessity for precise, adaptable solutions to address the diverse range of workplace scenarios. Furthermore, this research venture into neural network optimizations has yielded significant learnings. It showcased how adjustments in network parameters and loss functions directly influence the generation of optimized poses, underlining the delicate balance required to improve ergonomic assessments while maintaining the naturalness and intent of human movements. This process enabled a deeper comprehension of the intricate relationship between machine learning algorithms and ergonomic principles, contributing valuable insights to both fields.

6.1 Research Questions

This section revisits the research questions previously outlined in the thesis, focusing on optimizing ergonomics through neural networks. Each question is crafted to deepen the understanding of how machine learning can be effectively applied to ergonomic assessments, exploring optimal postures, the impact of neural optimizations, and ideal hyperparameter configurations for realistic and ergonomic results.

RQ1: What constitutes optimal ergonomics, and which metrics describe ergonomics?

Optimal ergonomics varies depending on specific tasks, individual characteristics, and workplace processes. Ergonomic assessment scores like RULA, REBA, OWAS, NIOSH Lifting Equation, and EWAS are each designed to address specific workplace needs and challenges. These tools focus on different aspects of ergonomic risk that are appropriate to the settings in which they are used. RULA is primarily used to evaluate risks affecting the upper body, including the neck, trunk, and upper limbs, making it ideal for environments with repetitive movements or static postures. REBA extends this assessment to the entire body, making it suitable for more dynamic settings, where movements involve the whole body, such as lifting patients. OWAS, meanwhile, categorizes working postures into risk levels, aiding industries that require physical labour. The NIOSH Lifting Equation offers a formulaic approach to reduce the physical demands of manual lifting by considering factors like the weight of the load and the position relative to the worker's body. This is particularly useful in industrial or warehouse environments. Similarly, EWAS provides a systematic method to evaluate ergonomic quality across various sectors, focusing on whole-body assessments. While distinct in focus, each tool serves the broader purpose of identifying high-risk areas and suggesting ergonomic improvements. The selection of a specific tool depends on the workplace context and the nature of tasks performed, ensuring that ergonomic interventions are both effective and appropriate to the setting.

However, specific movements, such as lifting, inherently carry higher ergonomic risks due to the nature of the action itself. While these movements can be optimized to a certain extent, there is a practical limit to how much their ergonomic risk can be reduced without compromising the intended purpose of the activity.

Automated ergonomic scoring systems are important in continuously refining workstations and processes. These assessments help identify specific ergonomic challenges, particularly in areas where ergonomics is inherently poor, such as frequent lifting tasks. It becomes possible to rethink workstations or processes strategically by pinpointing these problematic areas. This reevaluation aims to minimize the occurrence of high-risk movements, reducing ergonomic risks. Such optimizations ensure that essential tasks are performed more safely and comfortably, enhancing overall workplace ergonomics.

RQ2: To what extent can neural networks and gradient-based optimisation be employed to optimize posture while preserving the intended action?

Gradient-based optimization, mainly through neural networks, neural style optimization, and generative adversarial networks, can effectively optimize posture for ergonomic benefits while maintaining the intended action. This thesis has demonstrated that these techniques are adequate, using a balanced optimization strategy incorporating ergonomic, structural, and discriminator losses. The study has established that by integrating ergonomic loss with structural and discriminator losses, these methods can successfully optimize postures without compromising the original function of the movement.

Ergonomic loss is crucial for modifying postures to minimize the risk of strain or injury, yet its application alone may lead to unnatural or impractical poses if not appropriately balanced. Structural loss plays a key role here, ensuring that adjustments to the posture do not disrupt the biomechanical integrity or the smoothness of the movement over time. Furthermore, discriminator loss, utilized through generative adversarial networks, guarantees that the optimized poses remain realistic and human-like. This balance between ergonomic improvement and preserving a natural, functional posture is critical in ensuring that the optimized movements are safe and feasible for real-world applications.

Thus, this thesis's findings confirm that a multifaceted approach incorporating ergonomic, structural, and discriminator losses is not only feasible but also effective in enhancing posture while retaining the essential characteristics of the intended action. This approach allows for a comprehensive optimization that respects the dual objectives of health and performance efficacy.

Incorporating human action recognition could further enhance this approach by ensuring that the context and intent of movements are understood and preserved, facilitating both ergonomic improvements and the functional efficiency of actions.

RQ3: Which hyperparameter combination constitutes ergonomic optimization qualitatively?

Effective ergonomic optimization requires a balance of hyperparameters governing the ergonomic, structural, and discriminator losses. The coefficients $\alpha = 0.3$, $\beta = 0.4$, and $\gamma = 0.3$ were empirically determined to strike an effective balance between ergonomic viability, structural consistency across frames and the realism of generated human poses. Adjusting these hyperparameters is crucial, as they influence the optimization's focus, guiding the neural network to improve ergonomics while maintaining smooth transitions and realistic human-like characteristics. This combination was identified through rigorous testing, employing objective measures to evaluate the reduction in ergonomic risk scores alongside subjective assessments to ensure the retention of natural movement dynamics. Objective measurements indicated an average drop of 1.8 in ergonomic risk scores, accompanied by a reduction in body angles.

6.2 Limitations / Future Directions

Even though the optimization process in ergonomic assessments improves ergonomics by focusing on key angles, there are notable limitations and problems inherent to this approach. The primary limitation in the current ergonomic scoring system, notably in methods like RULA (Rapid Upper Limb Assessment), stems from an inadequate representation of all necessary angles by the keypoints provided by standard pose estimation frameworks. RULA, for instance, models the trunk as a single vertex. In contrast, in more detailed keypoint extraction frameworks, such as the one used in this study, the spine is represented by at least two vertices—illustrated by the inclusion of a midpoint at the spine, for example, keypoint 7 in Figure 6.1. This discrepancy can lead to significant inaccuracies during the optimization process.

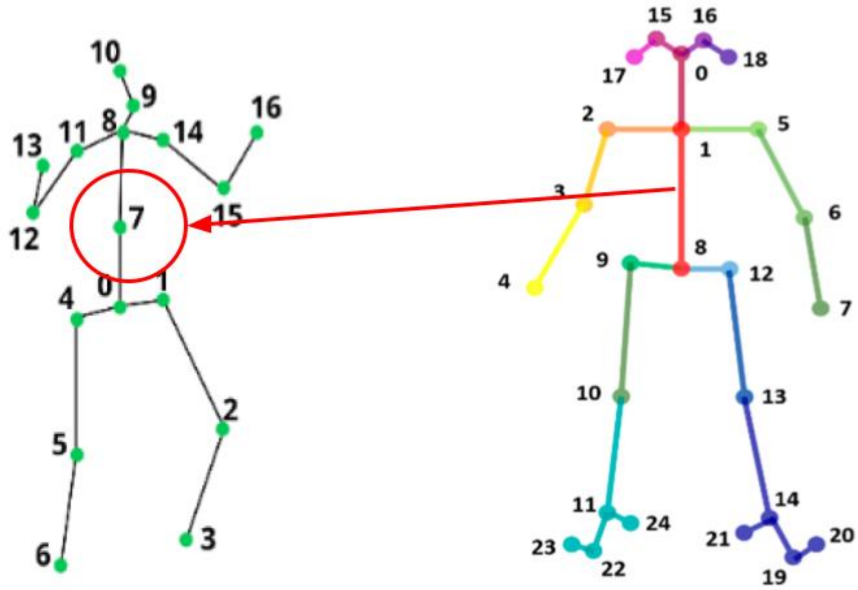


Figure 6.1: Comparison of ergonomic keypoint models: On the left is a detailed stick figure from this thesis with spine keypoint No. 7; on the right is a RULA model stick figure [KTK20] with a simplified single spine keypoint.

Specifically, while ergonomic scores like RULA focus on optimizing certain critical key angles, other keypoints that do not directly influence these scores can inadvertently shift. This shift occurs because the optimization process does not uniformly consider all body parts but instead focuses only on those angles directly contributing to the ergonomic score. As a result, optimizing the spine with two vertices in our approach, as opposed to one in RULA, can lead to incorrect alignment or positioning of the spine during pose recommendations. This issue highlights the need for a holistic approach to pose optimization, where the human body is treated as a cohesive unit rather than a collection of separate parts. Ensuring that changes in one area do not adversely affect the overall posture or lead to unergonomic positions. To address this, it is essential to modify existing frameworks on ergonomic scoring techniques to incorporate all angles into the calculation, allowing for a more comprehensive and effective optimization of poses. This would involve redefining key structures like the spine in keypoint extraction models to reflect more anatomical detail and functional relevance.

The next significant issue encountered in the optimisation process is that while the angles between keypoints are maintained, there can be a substantial alteration in the lengths of body parts. This results in generated poses that, although optimised for ergonomic angles, diverge from realistic human proportions. Such alterations might yield a figure with disproportionately elongated limbs or compressed torso segments, which, while theoretically ergonomic, are impractical and unnatural for actual physical tasks. Incorporating length constraints into the optimisation process could be essential to address the discrepancy in keypoint lengths after optimisation. We could maintain anatomical accuracy alongside ergonomic improvements by fixing the lengths of body parts to remain consistent with their pre-optimisation measurements. This approach would involve adjusting the optimisation algorithm to consider the natural lengths of limbs and other body segments, ensuring that they do not deviate from realistic human proportions.

A different approach to address this issue would be developing a more sophisticated GAN architecture that incorporates angle constraints between key points. Integrating angle constraints into the GAN's generator and discriminator components could substantially enhance the model's ability to generate anatomically coherent poses. For the generator, incorporating angle constraints would enable a more informed generation of poses by understanding human joints' natural range of motion and anatomical limitations. This leads to the production of poses that are not only ergonomically optimised but also resemble natural human movements more closely.

Similarly, for the discriminator, angle constraints would provide a robust criterion for evaluating the anatomical plausibility of generated poses. By assessing whether the angles between key points fall within realistic human limits, the discriminator could more effectively guide the generator towards producing ergonomically beneficial and anatomically plausible poses. This advancement in GAN architecture holds the potential to significantly improve the quality of pose generation, ensuring that the poses are ergonomically sound while maintaining the integrity of the intended movements.

Moreover, due to its lack of context awareness, the current optimization framework faces significant challenges in scenarios involving dynamic actions, such as lifting objects. The system's inability to perceive non-human elements, like the object being lifted, as well as the specific activities being performed, means that it optimizes the human pose in isolation. This oversight can result in recommendations for poses that are ergonomically optimized in theory but do not facilitate or may even hinder the actual physical task, such as lifting or moving an object. For instance, a pose that is deemed ideal for lifting based on spinal alignment and limb positioning might become impractical if it does not consider the height, weight, or positioning of the object to be lifted, as well as the dynamics of the human activity involved. These limitations underscore the need for a more advanced approach that can address these contextual challenges and enhance workplace safety and efficiency.

The lack of contextual understanding in the current optimization framework affects the practical application of ergonomic recommendations and risks the safety and efficiency of workplace tasks. It underscores the need for an advanced optimization approach that integrates the entire environment, considering both the objects being interacted with and the specific human activities. This could significantly enhance the utility and applicability of ergonomic assessments in dynamic and varied work settings. Incorporating human activity recognition into this framework could significantly address these limitations. By utilizing HAR to detect and analyze the specific activities and interactions within a workspace, ergonomic recommendations could be more accurately tailored to the user's context, ensuring that adjustments are both relevant and supportive of the task at hand. This integration fosters a more nuanced and effective ergonomic optimization, where recommendations genuinely reflect and enhance workplace safety and efficiency.

CHAPTER 7

Conclusion

In conclusion, this thesis has laid the foundation for integrating gradient-based optimization technologies in ergonomic assessments, demonstrating a potential for improving workplace health and safety. The novel use of neural style transfer and generative adversarial networks in optimizing human postures represents a promising step towards automating and refining ergonomic practices. Neural style transfer, in this context, involves adapting the stylistic elements of ideal ergonomic postures to real-world human movements captured in various workplace settings.

The optimization process is inspired by neural style transfer, integrating a composite loss function that includes ergonomic, structural, and discriminator losses. The ergonomic loss is specifically tailored to optimize postures for better ergonomic compliance. Structural loss ensures consistency and coherence between frames, maintaining the structural integrity of the sequence of movements. The discriminator loss, sourced from a generative adversarial network, critically assesses whether the optimized poses remain realistic and human-like. Together, these components facilitate a balanced optimization that improves posture ergonomics and preserves the natural flow and structure of movements across frames.

Expanding on this, integrating these three distinct loss functions allows for a robust and multifaceted approach to ergonomic optimization. The ergonomic loss component is essential as it directly targets the reduction of physical strain and the enhancement of comfort by adjusting key postural angles and alignments based on established ergonomic criteria. This is particularly important in tasks that require repetitive movements or long durations of static posture, where poor ergonomics can lead to significant health issues such as musculoskeletal disorders.

The structural loss also plays an important role in ensuring that the transitions between poses are smooth and realistic, which is essential not only for the visual coherence of the sequence but also for the practical implementation of these poses in real-world tasks. By maintaining structural integrity, the optimization process prevents abrupt or unnatural movements that could disrupt the activity flow.

Lastly, the discriminator loss, a key component of the GAN, provides a critical feedback mechanism that evaluates the authenticity of the optimized poses. By comparing them against the learned representation of human poses, the discriminator ensures that the modifications made during the optimization do not stray too far from natural human movements, which is vital for the practical applicability of the recommendations. This component helps refine the model continuously, enhancing its ability to produce ergonomically optimized poses that are realistic and applicable in various scenarios.

Potential future work involves refining these technologies to enhance their accuracy and applicability. This includes improving the learning algorithms to handle a more comprehensive array of human activities and postures, expanding the ergonomic and biomechanical models to encompass a broader spectrum of workplace scenarios, and integrating feedback mechanisms to allow the system to adapt based on user interaction and real-world testing. Additionally, exploring virtual and augmented reality integration could offer new ways to train individuals in ergonomic practices more effectively.

List of Figures

List of Tables

Bibliography

[Bev15] Stephen Bevan. Economic impact of musculoskeletal disorders (msds) on work in europe. *Best Practice & Research Clinical Rheumatology*, 29(3):356–373, 2015.

[BHSS07] S Bao, N Howard, P Spielholz, and B Silverstein. Two posture analysis approaches and their application in a modified rapid upper limb assessment evaluation. *Ergonomics*, 50(12):2118–2136, 2007.

[BW20] Xuqi Bai and Hendro Wicaksono. How relevant are environmental factors in the ergonomic performance assessments? *Procedia Manufacturing*, 52:325–330, 2020.

[Dav05] Geoffrey C David. Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders. *Occupational medicine*, 55(3):190–199, 2005.

[Den22] Tiancan Deng. A survey of convolutional neural networks for image classification: Models and datasets. In *2022 international conference on big data, information and computer network (BDICN)*, pages 746–749. IEEE, 2022.

[DKS24] Omar Drljevic, David Kostolani, and Sebastian Schlund. Call for action: Why ai in ergonomics does not translate to improved work postures. *GfA*, 2024.

[DMM05] Patrick G Dempsey, Raymond W McGorry, and Wayne S Maynard. A survey of tools and methods used by certified professional ergonomists. *Applied ergonomics*, 36(4):489–503, 2005.

[EMMV+22] Ilias El Makrini, Glenn Mathijssen, Sten Verhaegen, Tom Verstraten, and Bram Vanderborght. A virtual element-based postural optimization method for improved ergonomics during human-robot collaboration. *IEEE Transactions on Automation Science and Engineering*, 19(3):1772–1783, 2022.

[FRM+22] Manlio Massiris Fernández, Sanjin Radoš, Krešimir Matković, M Eduard Gröller, and Claudio Delrieux. Ergoexplorer: Interactive ergonomic risk assessment from video collections. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):43–52, 2022.

[FRS23] Clara Fischer, Patrick Rupprecht, and Sebastian Schlund. Different approaches of conducting ergonomic assessment utilizing digital human models and motion capture in industrial site assembly. 2023.

[GA21] Miniar Ben Gamra and Moulay A Akhloufi. A review of deep learning techniques for 2d and 3d human pose estimation. *Image and Vision Computing*, 114:104282, 2021.

[GEB15] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576*, 2015.

[GPAM+20] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2020.

[HM00] Sue Hignett and Lynn McAtamney. Rapid entire body assessment (reba). *Applied ergonomics*, 31(2):201–205, 2000.

[Hur88] Clifford M. Hurvich. A mean squared error criterion for time series data windows. *Biometrika*, 75(3):485–490, 1988.

[ITDÁ23] Guillermo Iglesias, Edgar Talavera, and Alberto Díaz-Álvarez. A survey on gans for computer vision: Recent research, analysis and taxonomy. *Computer Science Review*, 48:100553, 2023.

[IZZE17] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks, 2017.

[JB23] Jyoti Jangade and Kanojia Sindhuben Babulal. Study on deep learning models for human pose estimation and its real time application. In *2023 6th International Conference on Information Systems and Computer Networks (ISCON)*, pages 1–6. IEEE, 2023.

[JCDG] Richard Graveling Joanne Crawford, Alice Davis, and Eva Giagloglou. Work-related musculoskeletal disorders: from research to practice.what can be learnt? `https://euosha.blumm.it/bundles/app/assets/website/download/Work_related_musculoskeletal_disorders_from_research_to_practice_What`. Accessed: 2023-11-8.

[JYF+19] Yongcheng Jing, Yezhou Yang, Zunlei Feng, Jingwen Ye, Yizhou Yu, and Mingli Song. Neural style transfer: A review. *IEEE transactions on visualization and computer graphics*, 26(11):3365–3385, 2019.

[JZ19] Licheng Jiao and Jin Zhao. A survey on the new generation of deep learning in image processing. *Ieee Access*, 7:172231–172263, 2019.

[KB14] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[KBH17] Somnath Kolgiri, Sheelratan Bansode, and Rahul Hiremath. A literature review on work-related body-part discomfort with respect of postures and movements. *IAETSD Journal for advanced research in applied sciences*, 4(5), 2017.

[KKK77] Osmo Karhu, Pekka Kansi, and Iikka Kuorinka. Correcting working postures in industry: A practical method for analysis. *Applied ergonomics*, 8(4):199–201, 1977.

[KKS+23] Artur Karimov, Ekaterina Kopets, Tatiana Shpilevaya, Evgenii Katser, Sergey Leonov, and Denis Butusov. Comparing neural style transfer and gradient-based algorithms in brushstroke rendering tasks. *Mathematics*, 11(10):2255, 2023.

[KSS+21] Woojoo Kim, Jaeho Sung, Daniel Saakes, Chunxi Huang, and Shuping Xiong. Ergonomic postural assessment using a new open-source human pose estimation technology (openpose). *International Journal of Industrial Ergonomics*, 84:103164, 2021.

[KTK20] Yuhi Kaihoko, Phan Xuan Tan, and Eiji Kamioka. Prevention of unintended appearance in photos based on human behavior analysis. *Information*, 11(10):468, 2020.

[KWS22] David Kostolani, Michael Wollendorfer, and Sebastian Schlund. Ergomaps: Towards interpretable and accessible automated ergonomic analysis. In *2022 IEEE 3rd International Conference on Human-Machine Systems (ICHMS)*, pages 1–7. IEEE, 2022.

[LGBD+14] Lars Lidgren, E Gomez-Barrena, GN Duda, W Puhl, and A Carr. European musculoskeletal health and mobility in horizon 2020: Setting priorities for musculoskeletal research and innovation. *Bone & joint research*, 3(3):48–50, 2014.

[LLRL21a] Sujee Lee, Li Liu, Robert Radwin, and Jingshan Li. Machine learning in manufacturing ergonomics: Recent advances, challenges, and opportunities. *IEEE Robotics and Automation Letters*, 6(3):5745–5752, 2021.

[LLRL21b] Sujee Lee, Li Liu, Robert Radwin, and Jingshan Li. Machine learning in manufacturing ergonomics: Recent advances, challenges, and opportunities. *IEEE Robotics and Automation Letters*, 6(3):5745–5752, 2021.

[MBI18] Khalid M Malik, Rena Beckerly, and Farnad Imani. Musculoskeletal disorders a universal source of pain and disability misunderstood and mismanaged: A critical analysis based on the us model of care. *Anesthesiology and pain medicine*, 8(6), 2018.

[MC] Benjamin Bittschi Mayrhuber Christine. Fehlzeitenreport 2021. krankheits- und unfallbedingte fehlzeiten in Österreich – frühintervention, wiedereingliederung und mentale gesundheit.

[MC93] Lynn McAtamney and E Nigel Corlett. Rula: a survey method for the investigation of work-related upper limb disorders. *Applied ergonomics*, 24(2):91–99, 1993.

[MGHF+21] Christian Maurer-Grubinger, Fabian Holzgreve, Laura Fraeulin, Werner Betz, Christina Erbe, Doerthe Brueggmann, Eileen M Wanke, Albert Nienhaus, David A Groneberg, and Daniela Ohlendorf. Combining ergonomic risk assessment (rula) with inertial motion capture technology in dentistry—using the benefits from two worlds. *Sensors*, 21(12):4077, 2021.

[MJW+20] Tewodros Legesse Munea, Yalew Zelalem Jembre, Halefom Tekle Weldegebriel, Longbiao Chen, Chenxi Huang, and Chenhui Yang. The progress of human pose estimation: A survey and taxonomy of models applied in 2d human pose estimation. *IEEE Access*, 8:133330–133348, 2020.

[Nie15] Michael A Nielsen. Neural networks and deep learning, 2015.

[NK21] Gourav Kumar Nayak and Eunsik Kim. Development of a fully automated rula assessment system based on computer vision. *International Journal of Industrial Ergonomics*, 86:103218, 2021.

[OLC+23] Yewei Ouyang, Ming Liu, Cheng Cheng, Yuchen Yang, Shiyi He, and Lan Zheng. Monitoring inattention in construction workers caused by physical fatigue using electrocardiograph (ecg) and galvanic skin response (gsr) sensors. *Sensors*, 23(17):7405, 2023.

[Par00] Ken C Parsons. Environmental ergonomics: a review of principles, methods and models. *Applied ergonomics*, 31(6):581–594, 2000.

[PB21] Behnoosh Parsa and Ashis G Banerjee. A multi-task learning approach for human activity segmentation and ergonomics risk assessment. pages 2352–2362, 2021.

[PBL24] Francesco Pistolesi, Michele Baldassini, and Beatrice Lazzerini. A human-centric system combining smartwatch and lidar data to assess the risk of musculoskeletal disorders and improve ergonomics of industry 5.0 manufacturing workers. *Computers in Industry*, 155:104042, 2024.

[PFGA19] Dario Pavllo, Christoph Feichtenhofer, David Grangier, and Michael Auli. 3d human pose estimation in video with temporal convolutions and semi-supervised training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7753–7762. IEEE, 2019.

[PKKC22] Prabesh Paudel, Young-Jin Kwon, Do-Hyun Kim, and Kyoung-Ho Choi. Industrial ergonomics risk analysis based on 3d-human pose estimation. *Electronics*, 11(20):3403, 2022.

[PSH+19] Behnoosh Parsa, Ekta U Samani, Rose Hendrix, Cameron Devine, Shashi M Singh, Santosh Devasia, and Ashis G Banerjee. Toward ergonomic risk prediction via segmentation of indoor object manipulation actions using spatiotemporal convolutional networks. *IEEE Robotics and Automation Letters*, 4(4):3153–3160, 2019.

[Rud16] Sebastian Ruder. An overview of gradient descent optimization algorithms, 2016.

[SCBB13] Karlheinz Schaub, Gabriele Caragnano, Bernd Britzke, and Ralph Bruder. The european assembly worksheet. *Theoretical Issues in Ergonomics Science*, 14(6):616–639, 2013.

[SCZZ19] Shiliang Sun, Zehui Cao, Han Zhu, and Jing Zhao. A survey of optimization methods from a machine learning perspective, 2019.

[Shu23] David Shulman. Optimization methods in deep learning: A comprehensive overview. February 2023.

[SJJ+21] Akhil Singh, Vaibhav Jaiswal, Gaurav Joshi, Adith Sanjeeve, Shilpa Gite, and Ketan Kotecha. Neural style transfer: A critical review. *IEEE Access*, 9:131583–131613, 2021.

[SSLS19] Aliaksandr Siarohin, Enver Sangineto, Stéphane Lathuilière, and Nicu Sebe. Appearance and pose-conditioned human image generation using deformable gans. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(4):1156–1171, 2019.

[Str17] Helmut Strasser. The “art of aging” from an ergonomics viewpoint–wisdoms on age. *Occupational Ergonomics*, 13(S1):1–24, 2017.

[Sun19] Ruoyu Sun. Optimization for deep learning: theory and algorithms. *arXiv preprint arXiv:1912.08957*, 2019.

[TLS+17a] Kian Sek Tee, Eugene Low, Hashim Saim, Wan Nurshazwani Wan Zakaria, Safinaz Binti Mohd Khialdin, Hazlita Isa, MI Awad, and Chin Fhong Soon. A study on the ergonomic assessment in the workplace. 1883(1), 2017.

[TLS+17b] Kian Sek Tee, Eugene Low, Hashim Saim, Wan Nurshazwani Wan Zakaria, Safinaz Binti Mohd Khialdin, Hazlita Isa, MI Awad, and Chin Fhong Soon. A study on the ergonomic assessment in the workplace. In *AIP conference proceedings*, volume 1883. AIP Publishing, 2017.

[WPAG94] Thomas R Waters, Vern Putz-Anderson, and Arun Garg. Applications manual for the revised niosh lifting equation, 1994.

[WSH+19] Yanyun Wang, Chunfeng Song, Yan Huang, Zhenyu Wang, and Liang Wang. Learning view invariant gait features with two-stream gan. *Neurocomputing*, 339:245–254, 2019.

[XAS+20] Wei Xue, Hong Ai, Tianyu Sun, Chunfeng Song, Yan Huang, and Liang Wang. Frame-gan: Increasing the frame rate of gait videos with generative adversarial networks. *Neurocomputing*, 380:95–104, 2020.

[ZO21] Junqi Zhao and Esther Obonyo. Applying incremental deep neural networks-based posture recognition model for ergonomics risk assessment in construction. *Advanced Engineering Informatics*, 50:101374, 2021.