Kinematic Equine Gait Analysis and Architecture Trait Phenotyping of Loblolly Pine Using 3D Stereo Machine Vision and Deep Learning

by

Nariman Niknejad

A thesis submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Auburn, Alabama
December 10, 2022

Keywords: 3D Reconstruction, Deep Learning, Stereo Matching, Equine Gait Analysis, Animal Pose Estimation, Loblolly Pine, Tree Architecture

Approved by

Dr. Yin Bao, Chair, Assistant Professor, Department of Biosystems Engineering
Dr. Jingyi Zheng, Assistant Professor, Department of Mathematics and Statistics
Dr. Tanzeel Rehman, Assistant Professor, Departments of Biosystems Engineering / Horticulture

Abstract

As agricultural technologies have progressed rapidly over the past decades, animal and plant phenotyping has become one of the primary research topics. Conventional phenotyping relies heavily on manual measurements, which are labor-intensive, time-consuming, and subject to error. The purpose of this thesis is to present an innovative solution for estimating animal pose and quantifying pine tree architecture traits using 3D stereo machine vision and deep learning techniques.

It is currently necessary to undergo a complex, costly, and labor-intensive procedure for equine locomotion research to conduct horse kinematic gait analysis (EKGA). For the measurement of equine biomechanical parameters, an automated stereo video processing pipeline has been developed and evaluated. DeepLabCut (DLC) was trained on stereo videos of 40 walking horses to detect body landmarks. Landmark detection was conducted using an ARIMA filter, which had RMSE and MAE values of 5.14 pixels and 4.87 pixels, respectively. An analysis of stride length (SL) and stance duration (SD) was performed as a case study. The Faster R-CNN model and the mode filter were applied to perform individual hoof gait phase detection, yielding precision and recall values of 0.83 and 0.95, respectively. A semi-global block matching algorithm (SGBM) was used to estimate the depth maps, and accuracy was assessed by comparing estimated head lengths to measurements taken in the field. Bland-Altman analysis for DLC-detected head coordinates when combined with SGBM, yielded a bias of -0.014m with upper and lower limits of agreement (LoA) of 0.03 m and -0.061m, in that order. Moreover, Bland-Altman analysis on SD and SL also revealed biases of -0.02 s and -0.042 m compared to image-level manual measurements. Furthermore, both the upper and lower LOAs for SD were 0.01907 and -0.24 seconds, and for SL

they were 0.04 and -0.12 meters. In summary, the proposed method shows promising potential for performing EKGA in an automated, cost-efficient, and rapid manner.

A similar approach was used to provide a high throughput phenotyping solution for pine trees in 3D. Loblolly pines have long been one of the most significant forest trees for producing saw-wood in the Southern United States. The yield potential of a pine tree is significantly impacted by its stem and branch characteristics. A low-throughput technique, visual grading, is currently used in progeny trials to measure economically significant features such as stem straightness, branch angle, and branch diameter. To phenotype pine architecture, stereo 3D imaging and deep learning were combined. The stem diameter, branch angle, and branch diameter of ten loblolly pine trees belonging to different families were measured manually in a progeny test. An annotated dataset was created using contour polygons on the branches and trunks of each tree of interest. To segment branches and trunks using the dataset, a pre-trained Mask R-CNN model was fine-tuned and tested. The semi-global block matching (SGBM) algorithm was employed to reconstruct the 3D shapes of small trunks and thin branches. After the 3D point clouds were extracted, they were further processed using principal component analysis (PCA), random sample consensus (RANSAC), and statistical outlier removal. As compared to manual measurements, the three system-derived parameters had RMSEs of 0.05 m, 5.0 degrees, and 5.6 mm. Bland-Altman analysis showed that stem diameter, branch angle, and branch diameter all had standard deviations of 0.005 m, 5.0 deg, and 5.4 mm, and biases of 0.011 m, -0.4 deg, and -1.4 mm, respectively. As a precision phenotyping tool for the characterization of loblolly pine tree architecture, the proposed system shows promising potential. By facilitating the selection of tree architecture that is highly productive and resilient to severe weather events and climatic variability, it facilitates a better understanding of tree architecture.

Acknowledgments

My journey at Auburn University, the Department of Biosystems Engineering, began in January 2020, during a global pandemic called Covid-19. My experience with pure research was limited, let alone with the exciting world of phenotyping and precision agriculture for smart agricultural applications. As of now, I believe I have achieved maturity and have made significant progress, both personally and professionally. These achievements would not have been possible without the support and inspiration of several key individuals.

To begin with, I would like to extend my deepest appreciation to my major advisor, Dr. Yin Bao. He provided me with the necessary materials and tools for the projects, enlightened me when I was unsure, and encouraged me to push the boundaries of innovation. My ongoing motivation to conduct groundbreaking research has been sparked by his passion, professionalism, creativity, and persistent pursuit of novel solutions in the field of smart agriculture. This is especially true when it comes to the use cases of 3D phenotyping. His kind supervision has provided me with every opportunity to develop my research interests, and without him, this thesis would not have been possible. Besides his stellar academic credentials, Dr. Bao is also very supportive and kind. He offered me rides to work and the lab when I did not have a car. My sincere gratitude reaches out to you for all the interesting conversations we've had over the last two years, both academic and personal.

It was my pleasure to receive constructive suggestions from my other committee members, Dr. Jingyi Zheng and Dr. Tanzeel Rehman. As well as their support, I have received assistance in every way from them. Besides being an expert in data science, Dr. Zheng is also an expert in machine learning. Research into modeling and image processing for plant phenotyping has been a major

focus of Dr. Rehman's career. Their work has inspired me to work on interdisciplinary problems and I have gained a more profound appreciation for the opportunities I have been given. Through Biosystems Engineering, I met these amazing people that I never would have encountered without it. Furthermore, I want to thank Dr. Ann Staiger who generously provided me with the knowledge I did not possess when I was working with horses. As well as her dedication and expertise in equine gait analysis, she was a significant help in the collection of data. Without her and Jessica Caro, her master's student, the equine gait analysis project would not have been feasible. I am grateful for the opportunity to have met them and to have worked with them.

I would like to extend my gratitude to my colleagues in the AgCypher Lab, Rafael Bidese Puhl and his lovely wife, Ana Gabriela Itokazu, Mary Beth Cassity, and Sharif Shabani. My fellow officemates Preetika Kaur, Vishawjot Singh, and Bijoy Chandra also supported me during the hard times. My research and life have been helped by them in many ways. The work and conversations I have had with these smart and talented people have been a pleasure.

Last but not least, I want to take this opportunity to thank my parents, Maziar Niknejad and Dr. Sepideh Ahmadnia, as well as my smart sister, Anahita Niknejad. Their unconditional love and support have always been the foundation of my success.

Table of Contents

List of Figures

List of Tables

CHAPTER 1. GENERAL INTRODUCTION

1.1 BACKGROUND

This thesis aims to solve two phenotyping problems, kinematic equine gait analysis and extracting pine architecture traits using a combination of deep learning models and 3D machine vision systems and algorithms. Thus, it consists of two parts. The first, explains our approach to extract the biomechanical properties of horses using the aforementioned tools. Following that, the same tools but different algorithms were used to perform a novel and high-throughput plant phenotyping pipeline for loblolly pine trees.

The following provides a background for the equine gait analysis section of the thesis. According to estimates, the horse industry contributes approximately 122 billion dollars to the U.S. economy each year (Grice, 2018). Moreover, horses do not only contribute to people's livelihoods in the U.S., but they also play a significant role in the agriculture, construction, tourism, mining, and public transportation sectors in developing countries (Behnke and Nakirya, 2012). In light of these economic and cultural implications, it is imperative to study biomechanical parameters. Equine performance is influenced by the quality of their locomotion whether they are used for racing, pleasure, or sport (Serra Bragança et al., 2018a). Horse owners can experience financial losses due to lameness and a possible end to their horse's athletic career. This is because their horses are unable to train or compete due to lameness. Lameness is not a disease, but merely a sign of a disturbance in locomotion (Dyson et al., 2008; Jeffcott et al., 1982). Lameness examinations are used to localize its root cause to determine possible veterinary treatment.

Analysis of equine gait patterns is crucial to improving the breeding of horses, predicting their performance potential, and reducing training costs (Barrey, 1999). In addition to determining the health of a horse's musculoskeletal system, this measurement can also be used to assess performance. Behavioral analysis and gait analysis have been shown to reduce the cost of training (Rose et al., 2009). Monitoring horse gait patterns regularly can assist in detecting injuries and lameness early. Gait analysis is also used to assess the effectiveness of training or the increase in fitness of the horse (Rose et al., 2009). Physiotherapy or another treatment may be necessary if the horse's gait patterns are abnormal or drastically different from normal. Lameness was a problem for 53-68% of Thoroughbred racehorses under four years of age in 1980 (Jeffcott et al., 1982). Based on the National Equine Health Survey of 2016, 33% of UK horses suffered from lameness (Shrestha et al., 2017). The biomechanical parameters of high-performing racehorses and sport horses are also analyzed using gait analysis (Echterhoff et al., 2018). The importance of equine gait analysis and its detection technique cannot be overstated.

Detailed background information is provided below for the phenotyping section of the thesis regarding loblolly pine trees. In the decades between 1930 and 1990, loblolly pines grew rapidly due to extensive plantings and natural regeneration. Loblolly pine (Pinus taeda L.) is the major timber species in the southeastern United States. Approximately 13.4 million hectares (45 percent) of southern U.S. commercial forest land were under this type of forest in 1989. Prior to the European settlement of North America, only 2 million predominantly loblolly pine forests occupied the southern United States. Humans have shaped the growth of the loblolly pine (Schultz, 1999). The Southern United States has greatly benefited from tree improvement programs and genetic engineering in forestry. The southeastern U.S. pine plantations have experienced increased growth for the past two decades thanks to the control of competing vegetation (Allen et al., 2005; Borders and Bailey, 2001; Wheeler et al., 2015; White et al., 2013). In forest genetics, a set of novel methodologies emerged about 1980 that allowed, for example, descriptions of individuals and populations based on DNA sequence information. Scientists rooted in basic plant biology and ecology have developed a largely distinct community around this work, one that is often removed from forestry practice. A decade of increasingly sophisticated efforts led to extensive plantations of improved loblolly pine. This was driven by a desire to increase productivity, disease resistance, seedling survival, and shorten harvest rotations, leading the Southern U.S. to become the world's most productive timber region (Allen et al., 2005). Pinus species have been studied for their genetic relationships between wood density and growth capacity. A moderately positive correlation has been found between height and Pinus taeda (Matziris and Zobel, 1973).

A tree's DBH (breast height diameter) measurement affects both forest management and progeny management procedures (silviculture treatments, thinning, and final cuttings). Commercial birch log harvesting from thinnings was initially done to furnish raw material for the manufacture of sound-knotted furniture, such as shelves and tables for the home (Kilpeläinen et al., 2011; Peuhkurinen et al., 2007; Uusitalo, 1997; Uusitalo and Isotalo, 2005). Tree branching geometry and the resulting crown shape greatly influence sunlight utilization as well as their knot footprint in the body of the stem of the trees. Tree crown shapes vary considerably among and within terrestrial biomes. A characteristic tree crown shape is nearly impossible to determine for a particular biome, partly due to the difficulty of describing tree crown shapes. This is because the shape varies depending on local growing conditions and the age of the tree. Most tropical and temperate forests consist of layers of tall, top-heavy trees, while the lower layers tend to be bushy. Many biophysical processes are indirectly impacted by branches, which are dominant factors controlling the exchange of matter between vegetation and the atmosphere, hydrosphere, and lithosphere (Li et al., 2002; Rosell et al., 2009; van der Zande et al., 2006). The majority of biomass in boreal forests comes from cones, but to describe pines, spruces, and larches as large cones would be too simplified. As observed from satellite data, most terrestrial forests consist mostly of top-heavy trees, i.e., biomass near the ground is relatively small compared to the above ground (Forrester et al., 2018; Lindh et al., 2018; Niklas, 1982).

It is therefore essential to measure the characteristics of the stem, branches, and needles to determine the characteristics of a desirable genotype (Repola, 2009). There are variations between genotypes when it comes to stem and branch traits (Salonen et al., 1997).

1.2 RELATED WORK

Related work to the equine gait analysis section of the thesis is as follows. Based on the speed and pattern of footfall, a horse's locomotion can be classified into four major gaits. Walking is a four-beat gait that always has two or three feet in contact with the ground at once and is the slowest. Compared to the walk, the trot is faster and consists of two diagonal beats with two moments of suspension (all four feet are off the ground at the same time (Vinika Gupta, 2021).

Subjective visual examination is widely accepted as an effective method for assessing most gait events. However, observers are limited by the restrictions of the human eye's maximum temporal resolution, as well as their limited perception of asymmetry and memory. Furthermore, visual evaluations are also subject to substantial drawbacks, notably low inter-observer agreement and a challenge to document gait changes consistently and interchangeably (S. Dyson, 2014; Hammarberg et al., 2016; Hewetson et al., 2006; Serra Bragança et al., 2018b; Thomsen et al., 2010).

Equine gait measurement systems were once limited to sophisticated gait labs because of financial and practical constraints, but they are now more affordable and practical, making them a routine part of everyday clinical practice. Consequently, this development raises new concerns about the validity and usefulness of available systems. During a given period of time, EKGA examines the change in position of body segments on a plane. Time, displacement, velocity, and acceleration of motion are described quantitatively by linear and angular variables (Barrey, 1999). By extracting the precise positions of body landmarks, gait parameters can be generated to objectively evaluate locomotion. The most developed motion-capturing system in EKGA is based on marker-based optical motion capture (OMC). A horse's upper body is tracked optically through image processing in products such as QHorse (Qualisys AB, Sweden). Static strain gauges can also be used to measure joint angles and displacements (Serra Bragança et al., 2018). An alternative way to evaluate gait is by using inertial measurement units (IMUs) (Serra Bragança et al., 2020) like Equinosis (Equinosis LLC, USA) (Bosch et al., 2018). It has been reported that optoelectronic and IMU-based motion systems can be used to evaluate equine hoof movement. It appears that the IMU sensors are more accurate when analyzing fast motions, such as landing durations and breakover durations (Hagen et al., 2021). The position of the sensors on a horse, however, affected the kinematic data obtained from EKGA systems equipped with inertial sensors. A sensor-based method is also limited by the number and position of markers that are determined beforehand, and therefore, cannot be altered after data collection (Moorman et al., 2012).

Since computer vision and deep learning have advanced rapidly in recent years, markerless motion capture technology has improved substantially. The development of deep learning-based video analytics tools has led to the development of markerless animal pose estimation tools. As an example, DeepLabCut (DLC) and LEAP pioneered the use of deep convolutional neural networks (CNNs) to detect body landmarks from a single camera frame (Mathis et al., 2018; Pereira et al., 2019). With transfer-learning and data augmentation techniques, it was demonstrated that DLC could achieve near-human-level labeling accuracy on a small dataset (Mathis et al., 2018; Vonstad et al., 2020). An imaging system with multiple views can be used for markerless 3D pose estimation by capturing a scene from numerous cameras at different viewing angles and triangulating the body landmarks' coordinates from these views. There have been a number of deep learning-based software tools developed, including a 3D version of DLC, AniPose, DANNCE, FreiPose, and DeepFly3D (Bala et al., 2020; Dunn et al., 2021; Günel et al., 2019; Karashchuk, Rupp, et al., 2021; Karashchuk, Tuthill, et al., 2021; Marshall et al., 2021; Zimmermann et al., 2020). Even though multi-view imaging systems are highly accurate, they can be costly to manage and require precise calibration of the camera when setting up a new scene. In conjunction with pose estimation software tools, a single binocular stereo camera can provide RGB and depth images that offer a cost-effective and easy-to-use alternative to 3D animal motion capture. In field conditions, it is possible to relocate stereo cameras without calibration quickly.

Below is the relevant work to the phenotyping of loblolly pine trees presented in the thesis. Using tools such as tape measures, calipers, protractor, and hypsometers, it is possible to measure tree height, stem diameter, and branch characteristics (Fleck et al., 2011). As it is critical to measure tree attributes and characteristics accurately when conducting progeny trials and genotype selection (Næsset et al., 2004). It is true that direct measurement methods are the most reliable and accurate, however, they cannot be repeated for large fields. To address the aforementioned issue, allometric models are sometimes fitted to empirical samples based on basic tree metrics such as stem diameter (Dhôte et al., 2000; Gaffrey et al., 1998; Tarp-Johansen et al., 1997). Fitting a model to measurements to make an estimation necessitates a time-consuming procedure of model calibration and uncertainty estimation, which again, demands manual and human-level measurements (Gupta et al., 2005). In the literature, three main non-contact techniques are used, including sight-, range sensing-, and image-based methods. Sight-based methods are commonly used in forest inventories and progeny test since pine tree morphological characterization relies on expert visual assessment. Assessing quality attributes requires extensive fieldwork and is only possible in surveys based on samples (Kankare et al., 2014).

LiDAR (light detection and ranging) and terrestrial laser scanning (TLS) are the most widely employed methods for digitalizing tree architectures based on range sensing (Gorte et al., 2004; Kankare et al., 2014; Liang et al., 2016; Pfeifer et al., 2004; Srinivasan et al., 2014). In spite of the fact that LiDAR can be used to rebuild tree architecture (Tienaho et al., 2022), LiDAR data collection is time-consuming, has low spatial resolution, and is expensive (Wilkes et al., 2017). Further, TLS produces a significant amount of data that must be processed, requires specialized expertise, and is extremely costly to process. Therefore, using the aforementioned tools to perform a high-throughput pine tree architecture analysis in a progeny test field remains difficult because of the high density of pine trees and the occlusion caused by branches and needles.

Even though 2D methods are generally used in image-based phenotyping techniques, they cannot accurately represent three-dimensional (3D) quantities (Gibbs et al., 2018; Kaminuma et al., 2004). Studies have demonstrated that 3D modeling of plants provides more accuracy and robustness (Apelt et al., 2015). It is essential to use 3D imaging for the phenotyping of plants in order to track precise geometry (Ziamtsov and Navlakha, 2020). As well as providing quantitative descriptions, 3D imaging also permits measurements of additional traits not possible with 2D images (Paulus et al., 2014). There is, however, no research that employs image-based approaches to rebuild pine trees in the field. Meanwhile, recent developments in multi-view stereo vision (MVS) have been made for a variety of other plants. Stereo machine vision uses more than two images to reconstruct 3D canopy models using passive imaging. Nguyen et al. (2016) use this method to reconstruct a plant canopy's 3D structure by using information like height, width, volume, and leaf area. A canopy model has been successfully reconstructed with this method (Bao et al., 2019; Salas-Fernandez et al., 2017). To reconstruct 3D models, Klodt and Cremers (2015) employed a multiple-view imaging method. Two webcams and a time-of-flight (ToF) camera were combined by Li and Tang (2017) to develop a low-cost three-dimensional plant morphological trait analysis system for corn seedlings. By using a low-cost setup, this study shows that it is possible to reconstruct a 3D scene at high spatial resolution. Inspired by the aforementioned studies in high-throughput plant phenotyping and with the recent developments in commercial RGB-Depth cameras, a 3D reconstruction of a scene can be achieved with a one-time factory-calibrated commercial binocular camera (Paul et al., 2020).

Plants and fruits, including apples, oranges, and other fruits, have been segmented and counted through deep learning-based approaches (Chu et al., 2021; de Melo et al., 2020; Ganesh et al., 2019; Yikun Liu et al., 2021). As well, some studies have used convolutional neural networks to detect and segment sorghum leaves and stems (Baweja et al., 2018; Sodhi et al., 2017; Xiang et al., 2021). Deep learning models have been successfully used to segment plant organs, such as the work done by Zhang et al. (2017), which identifies branches in apple trees using R-CNN networks.

1.3 RESEARCH OBJECTIVES

The purpose of the first study was to investigate the feasibility of combining stereo 3D machine vision and deep convolutional neural networks (CNNs) to automate EKGA in the field. The specific research objectives were to develop and evaluate a method for tracking body landmarks in 3D space and measuring stride lengths and stance durations from side-viewing stereo videos of horses in motion. This pipeline was designed to produce SL and SD as outputs because they play a crucial role in validating hypotheses in EKGA and lameness detection research. In the future, we will be able to apply the developed pipeline to studies focusing on equine locomotion, such as gait and sports performance, in which phenotyping of biomechanical parameters is paramount.

The second objective in this thesis was to examine whether it would be feasible to quantify branch and stem characteristics of loblolly pine trees using low-cost and non-contact stereo machine vision, deep learning, and image and point cloud processing methods. The specific research objectives were to 1) develop an image processing pipeline to measure stem diameter, branch angle and branch diameter, and 2) evaluate the performance of the pipeline in measuring the aforementioned traits at the different height levels along the tree stem.

1.4 THESIS OUTLINE

The thesis consists of two journal articles that describe our experiments combining deep learning models and stereo cameras: (1) equine kinematic gait analysis using stereo videography and deep learning: stride length and stance duration estimation, and (2) phenotyping of architecture traits of loblolly pine trees using stereo machine vision and deep learning models: stem diameter, branch angle, and branch diameter. The introduction and objectives of the study are presented in Chapter 1. The first article (Chapter 2) describes the study in which our developed system was used to identify and track horse body landmarks as well as measure stride length and stance duration. In the second article (Chapter 3), a detailed description of the developed system for the 3D reconstruction of loblolly pine trees as well as the measurement of their traits using point cloud analysis tools was presented. A general conclusion and a list of recommendations for future research are drawn in Chapter 4.

1.5 REFERENCE

Allen, H. L., Fox, T. R., Campbell, R. G. (2005). What is ahead for intensive pine plantation silviculture in the south? *Southern Journal of Applied Forestry*, *29*(2), 62–69. https://doi.org/10.1093/SJAF/29.2.62

Bala, P. C., Eisenreich, B. R., Bum, S., Yoo, M., Hayden, B. Y., Park, H. S., Zimmermann, J., and Hayden, B. (2020). OpenMonkeyStudio: Automated Markerless Pose Estimation in Freely Moving Macaques. BioRxiv, 2020.01.31.928861. https://doi.org/10.1101/2020.01.31.928861

Bao, Y., Tang, L., Srinivasan, S., Schnable, P. S. (2019). Field-based architectural traits characterisation of maize plant using time-of-flight 3D imaging. *Biosystems Engineering*, *178*, 86–101. https://doi.org/10.1016/J.BIOSYSTEMSENG.2018.11.005

Barrey, E. (1999). Methods, applications and limitations of gait analysis in horses. *The Veterinary Journal*, *157*(1), 7–22. https://doi.org/10.1053/TVJL.1998.0297

Baweja, H. S., Parhar, T., Mirbod, O., Nuske, S. (2018). StalkNet: A deep learning pipeline for high-throughput measurement of plant stalk count and stalk width. *Springer Proceedings in Advanced Robotics*, *5*, 271–284. https://doi.org/10.1007/978-3-319-67361-5_18/COVER

Behnke, R., Nakirya, M. (2012). The contribution of livestock to the Ugandan economy. *IGAD Centre for Pastoral Areas and Livestock Development*, (02).

Bentley, L. P., Stegen, J. C., Savage, V. M., Smith, D. D., von Allmen, E. I., Sperry, J. S., Reich, P. B., and Enquist, B. J. (2013). An empirical assessment of tree branching networks and implications for plant allometric scaling models. *Ecology Letters*, *16*(8). https://doi.org/10.1111/ele.12127

Bernotas, G., Scorza, L. C. T., Hansen, M. F., Hales, I. J., Halliday, K. J., Smith, L. N., Smith, M. L., and McCormick, A. J. (2019). A photometric stereo-based 3D imaging system using computer vision and deep learning for tracking plant growth. *GigaScience*, *8*(5). https://doi.org/10.1093/GIGASCIENCE/GIZ056

Borders, B. E., Bailey, R. L. (2001). Loblolly Pine—Pushing the Limits of Growth. *Southern Journal of Applied Forestry*, *25*(2), 69–74. https://doi.org/10.1093/SJAF/25.2.69

Bosch, S., Serra Bragança, F., Marin-Perianu, M., Marin-Perianu, R., van der Zwaag, B. J., Voskamp, J., Back, W., van Weeren, R., and Havinga, P. (2018). EquiMoves: A wireless networked inertial measurement system for objective examination of horse gait. *Sensors*, *18*(3), 850. https://doi.org/10.3390/s18030850

Chu, P., Li, Z., Lammers, K., Lu, R., Liu, X. (2021). Deep learning-based apple detection using a suppression mask R-CNN. *Pattern Recognition Letters*, *147*, 206–211. https://doi.org/10.1016/J.PATREC.2021.04.022

de Melo, R. F., de Lima, G. L., Corrêa, G. R., Zatt, B., de Aguiar, M. S., Nachtigall, G. R., Araújo, R. M. (2020). Diagnosis of apple fruit diseases in the wild with Mask R-CNN. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12319 LNAI*, 256–270. https://doi.org/10.1007/978-3-030-61377-8_18/COVER

Dhôte, J. F., Hatsch, E., Rittié, D. (2000). Stem taper curves, volume tables and volume yield compartments in Sessile Oak. *Annals of Forest Science*, *57*(2). https://doi.org/10.1051/forest:2000164

Dunn, T. W., Marshall, J. D., Severson, K. S., Aldarondo, D. E., Hildebrand, D. G. C., Chettih, S. N., … Ölveczky, B. P. (2021). Geometric deep learning enables 3D kinematic profiling across species and environments. *Nature Methods*, *18*(5), 564–573. https://doi.org/10.1038/S41592-021-01106-6

Dyson, P. K., Jackson, B. F., Pfeiffer, D. U., Price, J. S. (2008). Days lost from training by two- and three-year-old Thoroughbred horses: a survey of seven UK training yards. *Equine Veterinary Journal*, *40*(7), 650–657. https://doi.org/10.2746/042516408X363242

Dyson, S. (2014). Recognition of lameness: Man versus machine. *The Veterinary Journal*, *201*(3), 245–248. https://doi.org/10.1016/J.TVJL.2014.05.018

Echterhoff, J. M., Haladjian, J., Brugge, B. (2018). Gait and jump classification in modern equestrian sports. *Proceedings - International Symposium on Wearable Computers, ISWC*, 88–91. https://doi.org/10.1145/3267242.3267267

Fleck, S., Mölder, I., Jacob, M., Gebauer, T., Jungkunst, H. F., Leuschner, C. (2011). Comparison of conventional eight-point crown projections with LIDAR-based virtual crown projections in a temperate old-growth forest. *Annals of Forest Science*, *68*(7), 1173–1185. https://doi.org/10.1007/S13595-011-0067-1/FIGURES/11

Forrester, D. I., Ammer, C., Annighöfer, P. J., Barbeito, I., Bielak, K., Bravo-Oviedo, A., Coll, L., del Río, M., Drössler, L., Heym, M., Hurt, V., Löf, M., den Ouden, J., Pach, M., Pereira, M. G., Plaga, B. N. E., Ponette, Q., Skrzyszewski, J., Sterba, H., and Svoboda, M. (2018). Effects of crown architecture and stand structure on light absorption in mixed and monospecific Fagus sylvatica and Pinus sylvestris forests along a productivity and climate gradient through Europe. *Journal of Ecology*, *106*(2), 746–760. https://doi.org/10.1111/1365-2745.12803

Gaffrey, D., Sloboda, B., Matsumura, N. (1998). Representation of tree stem taper curves and their dynamic, using a linear model and the centroaffine transformation. *Journal of Forest Research 1998 3:2*, *3*(2), 67–74. https://doi.org/10.1007/BF02760304

Ganesh, P., Volle, K., Burks, T. F., Mehta, S. S. (2019). Deep Orange: Mask R-CNN based orange detection and segmentation. *IFAC-PapersOnLine*, *52*(30), 70–75. https://doi.org/10.1016/J.IFACOL.2019.12.499

Gazda, A., Kędra, K. (2017). Tree architecture description using a single-image photogrammetric method. *Dendrobiology*, *78*, 124–135. https://doi.org/10.12657/denbio.078.012

Gorte, B., Gorte, B., Pfeifer, N. (2004). Structuring laser-scanned trees using 3D mathematical morphology. *INTERNATIONAL ARCHIVES OF PHOTOGRAMMETRY AND REMOTE SENSING. 35(B5*, 929--933. Retrieved from http://130.203.136.95/viewdoc/summary?doi=10.1.1.638.5359

Grice, A. L. (2018). 2017 American Horse Council economic impact study. *Proceedings of the 64th Annual Convention of the American Association of Equine Practitioners, San Francisco, California, USA, 1-5 December 2018*, 502–504.

Günel, S., Rhodin, H., Morales, D., Campagnolo, J., Ramdya, P., Fua, P. (2019). Deepfly3D, a deep learning-based approach for 3D limb and appendage tracking in tethered, adult Drosophila. *ELife*, *8*. https://doi.org/10.7554/ELIFE.48571

Hagen, J., Jung, F. T., Brouwer, J., Bos, R. (2021). Detection of equine hoof motion by using a hoof-mounted inertial measurement unit sensor in comparison to examinations with an optoelectronic technique - A Pilot Study. *Journal of Equine Veterinary Science*, *101*. https://doi.org/10.1016/j.jevs.2021.103454

Hammarberg, M., Egenvall, A., Pfau, T., Rhodin, M. (2016). Rater agreement of visual lameness assessment in horses during lungeing. *Equine Veterinary Journal*, *48*(1), 78–82. https://doi.org/10.1111/EVJ.12385

Hannrup, B., Ekberg, I., Persson, A. (2010). Genetic correlations among wood, growth capacity and stem traits in pinus sylvestris. *Http://Dx.Doi.Org/10.1080/028275800750014966*, *15*(2), 161–170. https://doi.org/10.1080/028275800750014966

He, K., Gkioxari, G., Dollár, P., Girshick, R. (2017). Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(2), 386–397. https://doi.org/10.48550/arxiv.1703.06870

Hewetson, M., Christley, R. M., Hunt, I. D., Voute, L. C. (2006). Investigations of the reliability of observational gait analysis for the assessment of lameness in horses; Investigations of the reliability of observational gait analysis for the assessment of lameness in horses. https://doi.org/10.1136/vr.158.25.852

Jeffcott, L. B., Rossdale, P. D., Freestone, J., Frank, C. J., Towers-Clark, P. F. (1982a). An assessment of wastage in thoroughbred racing from conception to 4 years of age. *Equine Veterinary Journal*, *14*(3), 185–198. https://doi.org/10.1111/J.2042-3306.1982.TB02389.X

Kankare, V., Joensuu, M., Vauhkonen, J., Holopainen, M., Tanhuanpää, T., Vastaranta, M., Hyyppä, J., Hyyppä, H., Alho, P., Rikala, J., and Sipi, M. (2014). Estimation of the timber quality of Scots pine with terrestrial laser scanning. *Forests 2014, Vol. 5, Pages 1879-1895*, *5*(8), 1879–1895. https://doi.org/10.3390/F5081879

Karashchuk, P., Rupp, K. L., Dickinson, E. S., Walling-Bell, S., Sanders, E., Azim, E., Brunton, B. W., and Tuthill, J. C. (2021). Anipose: A toolkit for robust markerless 3D pose estimation. *Cell Reports*, *36*. https://doi.org/10.1016/j.celrep.2021.109730

Karashchuk, P., Tuthill, J. C., Brunton, B. W. (2021). The DANNCE of the rats: a new toolkit for 3D tracking of animal behavior. *Nature Methods*, *18*(5), 460–462. https://doi.org/10.1038/s41592-021-01110-w

Kilpeläinen, H., Lindblad, J., Heräjärvi, H., Verkasalo, E. (2011). Saw log recovery and stem quality of birch from thinnings in southern Finland. *Silva Fennica*, *45*(2). https://doi.org/10.14214/sf.117

Kolhar, S., Jagtap, J. (2021). Plant trait estimation and classification studies in plant phenotyping using machine vision – A review. *Information Processing in Agriculture*. https://doi.org/10.1016/j.inpa.2021.02.006

Liang, X., Kankare, V., Hyyppä, J., Wang, Y., Kukko, A., Haggrén, H., Yu, X., Kaartinen, H., Jaakkola, A., Guan, F., Holopainen, M., and Vastaranta, M. (2016). Terrestrial laser scanning in forest inventories. *ISPRS Journal of Photogrammetry and Remote Sensing*, *115*, 63–77. https://doi.org/10.1016/J.ISPRSJPRS.2016.01.006

Lindh, M., Falster, D. S., Zhang, L., Dieckmann, U., Brännström, Å. (2018). Latitudinal effects on crown shape evolution. *Ecology and Evolution*, *8*(16), 8149–8158. https://doi.org/10.1002/ECE3.4275

Liu, Y., Yang, G., Huang, Y., Yin, Y. (2021). SE-Mask R-CNN. *Journal of Intelligent and Fuzzy Systems*, *41*(6), 6715–6725. https://doi.org/10.3233/JIFS-210597

Marshall, J. D., Aldarondo, D. E., Dunn, T. W., Wang, W. L., Berman, G. J., Ölveczky, B. P. (2021). Continuous whole-body 3D kinematic recordings across the rodent behavioral repertoire. *Neuron*, *109*(3), 420-437.e8. https://doi.org/10.1016/J.NEURON.2020.11.016

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, *21*(9), 1281–1289. https://doi.org/10.1038/s41593-018-0209-y

Matziris, D. I., Zobel, B. J. (1973). Inheritance and correlations of juvenile characteristics in loblolly pine (Pinus taeda L.). *Silvae Genet*, *22*(1/2), 38–45.

Moorman, V. J., Reiser II, R. F., McIlwraith, C. W., Kawcak, C. E. (2012). Validation of an equine inertial measurement unit system in clinically normal horses during walking and trotting. *American Journal of Veterinary Research*, *73*(8), 1160–1170. https://doi.org/10.2460/ajvr.73.8.1160

Næsset, E., Gobakken, T., Holmgren, J., Hyyppä, H., Hyyppä, J., Maltamo, M., Nilsson, M., Olsson, H., Persson, Å., and Söderman, U. (2004). Laser scanning of forest resources: The Nordic experience. In *Scandinavian Journal of Forest Research* (Vol. 19). https://doi.org/10.1080/02827580410019553

Niklas, K. J. (1982). Computer simulations of early land plant branching morphologies: canalization of patterns during evolution? *Paleobiology*, *8*(3), 196–210. https://doi.org/10.1017/S009483730000693X

Paul, A., Ghosh, S., Das, A. K., Goswami, S., das Choudhury, S., Sen, S. (2020). A review on agricultural advancement based on computer vision and machine learning. *Advances in Intelligent Systems and Computing*, *937*, 567–581. https://doi.org/10.1007/978-981-13-7403-6_50/COVER

Pereira, T. D., Aldarondo, D. E., Willmore, L., Kislin, M., Wang, S. S.-H., Murthy, M., Shaevitz, J. W. (2019). Fast animal pose estimation using deep neural networks. *Nature Methods*, *16*(1), 117–125. https://doi.org/10.1038/s41592-018-0234-5

Peuhkurinen, J., Maltamo, M., Malinen, J., Pitkänen, J., Packalén, P. (2007). Preharvest measurement of marked stands using airborne laser scanning. *Forest Science*, *53*(6).

Pfeifer, N., Gorte, B., Winterhalder, D. (2004). Automatic reconstruction of single trees from terrestrial laser scanner data. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* (Vol. 35).

Phattaralerphong, J., Sinoquet, H. (2005). A method for 3D reconstruction of tree crown volume from photographs: Assessment with 3D-digitized plants. *Tree Physiology*, *25*(10), 1229–1242. https://doi.org/10.1093/TREEPHYS/25.10.1229

Polder, G., Hofstee, J. W. (2014). Phenotyping large tomato plants in the greenhouse using a 3D light-field camera. *American Society of Agricultural and Biological Engineers Annual International Meeting 2014, ASABE 2014*, *1*, 1-. https://doi.org/10.13031/AIM.20141882255

Repola, J. (2009). Biomass equations for Scots pine and Norway spruce in Finland. *Silva Fennica*, *43*(4). https://doi.org/10.14214/sf.184

Rose, N. S., Northrop, A. J., Brigden, C. v., Martin, J. H. (2009). Effects of a stretching regime on stride length and range of motion in equine trot. *The Veterinary Journal*, *181*(1), 53–55. https://doi.org/10.1016/j.tvjl.2009.03.010

Salas Fernandez, M. G., Bao, Y., Tang, L., Schnable, P. S. (2017). A high-throughput, field-based phenotyping technology for tall biomass crops. *Plant Physiology*, *174*(4), 2008–2022. https://doi.org/10.1104/PP.17.00707

Salonen, T., Oja, S., Julius Boutelje, B., Brskke, F. H., Czaplewski, R. L., Hyde, W. F., Mattson, W. J. (1997). Progeny trial estimates of genetic parameters for growth and quality traits in Scots pine. *Silva Fennica*, *31*(1), 3–12. https://doi.org/10.14214/SF.A8506

Schultz, R. P. (1999). Loblolly - The pine for the twenty-first century. In *New Forests* (Vol. 17). https://doi.org/10.1023/A:1006533212151

Serra Bragança, F. M., Broomé, S., Rhodin, M., Björnsdóttir, S., Gunnarsson, V., Voskamp, J. P., Persson-Sjodin, E., Back, W., Lindgren, G., Novoa-Bravo, M., Gmel, A. I., Roepstorff, C., van der Zwaag, B. J., Van Weeren, P. R., and Hernlund, E. (2020). Improving gait classification in horses by using inertial measurement unit (IMU) generated data and machine learning. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-73215-9

Serra Bragança, F. M., Rhodin, M., van Weeren, P. R. (2018). On the brink of daily clinical application of objective gait analysis: What evidence do we have so far from studies using an induced lameness model? *The Veterinary Journal*, *234*, 11–23. https://doi.org/10.1016/j.tvjl.2018.01.006

Shelhamer, E., Long, J., Darrell, T. (2014). Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(4), 640–651. https://doi.org/10.48550/arxiv.1411.4038

Shrestha, A., Bhattarai, D., Thapa, B., Basel, P., Wagle, R. R. (2017). Health care workers' knowledge, attitudes and practices on tuberculosis infection control, Nepal. *BMC Infectious Diseases*, *17*(1), 1–7. https://doi.org/10.1186/S12879-017-2828-4/TABLES/5

Sodhi, P., Vijayarangan, S., Wettergreen, D. (2017). In-field segmentation and identification of plant structures using 3D imaging. *IEEE International Conference on Intelligent Robots and Systems*, *2017-September*, 5180–5187. https://doi.org/10.1109/IROS.2017.8206407

Srinivasan, S., Popescu, S. C., Eriksson, M., Sheridan, R. D., Ku, N. W. (2014). Multi-temporal terrestrial laser scanning for modeling tree biomass change. *Forest Ecology and Management*, *318*, 304–317. https://doi.org/10.1016/j.foreco.2014.01.038

Tarp-Johansen, M. J., Skovsgaard, J. P., Madsen, S. F., Johannsen, V. K., Skovgaard, I. (1997). Compatible stem taper and stem volume functions for oak (Quercus robur L and Q petraea (Matt) Liebl) in Denmark. *Annales Des Sciences Forestières*, *54*(7), 577–595. https://doi.org/10.1051/FOREST:19970701

Taylor, E. L., Jack, S. B. (1998). *Using small-format, stereo photogrammetry to characterize changes in the three-dimensional crown architecture of loblolly pine (Pinus taeda L.). ProQuest Dissertations and Theses*. Ann Arbor. Retrieved from http://spot.lib.auburn.edu/login?url=https://www.proquest.com/dissertations-theses/using-small-format-stereo-photogrammetry/docview/304508366/se-2?accountid=8421

Thomsen, M. H., Persson, A. B., Jensen, A. T., Sørensen, H., Andersen, P. H. (2010). Agreement between accelerometric symmetry scores and clinical lameness scores during experimentally induced transient distension of the metacarpophalangeal joint in horses. *Equine Veterinary Journal*, *42*(SUPPL. 38), 510–515. https://doi.org/10.1111/J.2042-3306.2010.00287.X

Tienaho, N., Yrttimaa, T., Kankare, V., Vastaranta, M., Luoma, V., Honkavaara, E., Koivumäki, N., Huuskonen, S., Hynynen, J., Holopainen, M., Hyyppä, J., and Saarinen, N. (2022). Assessing Structural Complexity of Individual Scots Pine Trees by Comparing Terrestrial Laser Scanning and Photogrammetric Point Clouds. *Forests 2022, Vol. 13, Page 1305*, *13*(8), 1305. https://doi.org/10.3390/F13081305

Triki, A., Bouaziz, B., Gaikwad, J., Mahdi, W. (2021). Deep leaf: Mask R-CNN based leaf detection and segmentation from digitized herbarium specimen images. *Pattern Recognition Letters*, *150*, 76–83. https://doi.org/10.1016/J.PATREC.2021.07.003

Tuong Nguyen, T., Slaughter, D., Maloof, J., Sinha Thuy Tuong Nguyen, N., Slaughter, D. C., Maloof, J. N., Sinha, N. (2016). Plant phenotyping using multi-view stereo vision with structured lights. *Https://Doi.Org/10.1117/12.2229513*, *9866*(17), 22–30. https://doi.org/10.1117/12.2229513

Uusitalo, J. (1997). Pre-harvest measurement of pine stands for sawing production planning. *Acta Forestalia Fennica*, *259*. https://doi.org/10.14214/aff.7519

Uusitalo, Jori, Isotalo, J. (2005). Predicting knottiness of Pinus sylvestris for use in tree bucking procedures. *Scandinavian Journal of Forest Research*, *20*(6). https://doi.org/10.1080/02827580500407109

Vinika Gupta. (2021). *Equine gait analysis, body part tracking using DeepLabCut and Mask R-CNN and biomechanical parameter extraction.* Auburn University, Auburn.

Wheeler, N. C., Steiner, K. C., Schlarbaum, S. E., Neale, D. B. (2015). The evolution of forest genetics and tree improvement research in the United States. *Journal of Forestry*, *113*(5), 500–510. https://doi.org/10.5849/JOF.14-120

White, T., Davis, J., Gezan, S., Hulcr, J., Jokela, E., Kirst, M., Martin, T. A., Peter, G., Powell, G., and Smith, J. (2013). Breeding for value in a changing world: past achievements and future prospects. *New Forests 2013 45:3*, *45*(3), 301–309. https://doi.org/10.1007/S11056-013-9400-X

Zimmermann, C., Schneider, A., Alyahyay, M., Brox, T., Diester, I. (2020). FreiPose: A deep learning framework for precise animal motion capture in 3d spaces. https://doi.org/10.1101/2020.02.27.967620

CHAPTER 2. EQUINE KINEMATIC GAIT ANALYSIS USING STEREO VIDEOGRAPHY AND DEEP LEARNING: STRIDE LENGTH AND STANCE DURATION

Highlights

- Stereo machine vision and deep learning techniques were investigated for in-field equine kinematic gait analysis.
- The proposed pipeline tracks equine body landmarks in 3D space and estimates stride length and stance duration.
- The system provides a cost-effective, rapid, and easy-to-use tool for equine locomotion research.

2.1 ABSTRACT

Equine kinematic gait analysis (EKGA) currently requires a complicated, expensive, and labor-intensive procedure for equine locomotion research. An automated stereo video processing pipeline was developed and evaluated for measuring equine biomechanical parameters. Using stereo videos of 40 different walking horses, a DeepLabCut (DLC) model was trained to detect body landmarks in individual frames. With an ARIMA filter, the landmark detection had RMSE and MAE values of 5.14 pixels and 4.87 pixels, respectively. As a case study, methods were developed to extract stride length (SL) and stance duration (SD). Individual hoof gait phase detection was achieved using a fine-tuned Faster R-CNN model and a mode filter, yielding precision and recall values of 0.83 and 0.95, respectively. The semi-global block matching (SGBM) algorithm was used to estimate depth maps and the accuracy was assessed by comparing head length estimation with in-field measurements. A Bland-Altman analysis for DLC-detected head coordinates when combined with SGBM, yielded a bias of -0.014m with upper and lower limits of agreement (LoA) of 0.03 m and -0.061m, respectively. Furthermore, Bland-Altman analyses on SD and SL when compared to image-level manual measurements showed biases of -0.02 s and -0.042 m, respectively. Also, upper and lower LOAs were 0.01907s and -0.24 s for SD and 0.04 m and -0.12 m for SL. The proposed method shows promising potential in performing EKGA in an automated, cost-effective, and rapid manner.

Keywords. 3D Reconstruction, Animal Pose Estimation, Deep Learning, Equine Kinematic Gait Analysis, Stereo Matching.

2.2 INTRODUCTION

It is estimated that the total annual contribution of the horse industry to the U.S. economy is approximately 122 billion dollars according to (Grice, 2018). Moreover, horses' relevance is not specific to the U.S, as they contribute to people's livelihoods specifically in developing countries in a wide range of sectors, including but not limited to agriculture, construction, tourism, mining, and public transportation (Behnke and Nakirya, 2012). Due to such economic and cultural impacts, studying biomechanical parameters is of paramount importance. This can be helpful in the performance evaluation of horses and to determine the health of a horse's musculoskeletal system. It has been demonstrated that studying gait and behavioral analysis can potentially reduce training costs (Barrey et al., 1995; Rose et al., 2009). Horse training effectiveness and fitness evaluations can be done via qualitative or quantitative analysis of the biomechanical properties of the horses. A variety of methods have been proposed to perform an analysis of the motion of horses to measure temporal stride parameters. Typically, veterinarians evaluate a horse's musculoskeletal system through a clinical examination where an expert observes the horse walking and trotting "in hand" (led by a person) to find any visible asymmetries in its movements and to identify the localization of lameness (Pfau et al., 2016). However, this approach is subjective and is based on the examiner's technical knowledge (Keegan et al., 2000) which can lead to different diagnoses from different examiners. Hence, equine kinematic gait analysis (EKGA) technologies that are quantitative, objective, and automated are of great value to a wide range of applications.

EKGA examines the change in position of body segments on a given plane for a determined period. Quantitatively, motions are described by linear and angular variables relating to time, displacement, velocity, and acceleration (Barrey, 1999). One method to objectively evaluate locomotion lies in generating gait parameters based on extracting the precise positions of body landmarks. Marker-based optical motion capture (OMC) systems are by far the most developed motion-capturing methods in EKGA. In commercial marker-based products such as QHorse (Qualisys AB, Sweden), reflective markers on the upper body of a horse are tracked optically via image processing techniques. The angle and displacement of a joint can also be measured using strain gauges (Serra Bragança et al., 2018). Another approach is to use inertial measurement units (IMUs) (Serra Bragança et al., 2020) such as Equinosis (Equinosis LLC, USA) as a technique of gait evaluation (Bosch et al., 2018). A few reports have examined the comparability of optoelectronic and IMU-based motion systems to evaluate equine hoof motion. When analyzing fast motions like landing duration and break over duration, the two systems show less agreement while the IMU sensors seem to perform better (Hagen et al., 2021). However, sensor positions on a horse were found to affect the kinematic data derived from EKGA systems equipped with inertial sensors (Moorman et al., 2012). Another limitation of sensor-based methods is the number and placement of markers that need to be determined beforehand, and thus, unalterable after data collection.

In comparison, markerless motion capture technology has been significantly improved due to the rapid advancements in computer vision and deep learning in recent years. Various deep learning-based video analytics tools have been developed for markerless animal pose estimation. For instance, DeepLabCut (DLC) and LEAP were the first to exploit the use of deep convolutional neural networks (CNNs) for animal pose estimation by performing frame-based body landmark detection from a single camera (Mathis et al., 2018; Pereira et al., 2019). It was demonstrated that DLC could achieve near-human-level labeling accuracy on a small dataset thanks to transfer-learning and data augmentation techniques (Mathis et al., 2018; Vonstad et al., 2020). For markerless three-dimensional (3D) pose estimation, a multi-view imaging system is needed to capture a scene from many cameras at different viewing points and triangulate the 3D coordinates of the detected body landmarks. Various deep learning-based software tools have been developed including the 3D version of DLC, AniPose (Karashchuk, Rupp, et al., 2021), DANNCE (Dunn et al., 2021; Karashchuk, Tuthill, et al., 2021), OpenMonkeyStudio (Bala et al., 2020), FreiPose (Zimmermann et al., 2020), and DeepFly3D (Günel et al., 2019). While highly accurate, multi-view imaging systems are costly, difficult to manage, and require accurate camera calibration during setup for new scenes. A single binocular stereo camera that provides RGB and depth images may offer a cost-effective and easy-to-use alternative for 3D animal motion capture in conjunction with the pose estimation software tools. A single stereo camera can be easily relocated without calibration to accommodate constraints under field conditions.

This study investigated the feasibility of combing stereo 3D machine vision and deep convolutional neural networks (CNN) to automate EKGA under field conditions. The specific research objectives were to develop and evaluate the performance of a data processing pipeline that can track body landmarks in 3D space and measure stride length (SL) and stance duration (SD) from side-viewing stereo videos of horses in locomotion. As a case study, SL and SD were selected to be the outputs of this pipeline because of their paramount importance in verifying various hypotheses in EKGA and lameness detection studies (Arkell et al., 2006; Keegan, 2007). Our developed pipeline can then be applied to equine genetic studies focusing on locomotion, such as gait and sports performance, which require precise phenotyping of biomechanical parameters.

2.3 MATERIALS AND METHODS

2.3.1 *Stereo Video Data Collection*

The developed stereo video acquisition system consisted of a ZED2 stereo camera (StereoLabs, France), a Jetson Xavier NX embedded computer (NVIDIA, USA), and a Wi-Fi router (TP-Link, China) as shown in Figure 2-1a. A web-based user interface was developed using Robot Operating System 2 (DiLuoffo et al., 2018) for camera control and monitoring on mobile devices (i.e., a laptop). The system was put on a tripod at a height of 1.5 m off the ground. The stereo videos were saved in the SVO format (a proprietary file format created by StereoLabs) at 15 frames per second (FPS) and with a resolution of 2208 × 1242 pixels for both the left and right cameras of the ZED2. The angular field-of-view of the stereo camera was 110° × 70°, horizontal × vertical, respectively. Each horse was handled to move from the left side to the right side of the stereo video frame. The distance between each horse in motion and the camera was maintained at approximately 3m. The duration of the videos varied between 65 and 85 frames. The horses included different breeds (i.e., Warmbloods, Morgans, Quarter Horses, and Thoroughbreds), and different coat colors. Prior to video recording, a single individual collected body measurements of each horse as outlined in (Brooks et al., 2010). Additionally, a variety of outdoor lighting conditions (i.e., sunny, cloudy, overcast, backlit, etc.) were present in the dataset. A subset of the collected left-view images of the stereo videos in the dataset is shown in Figure 2-1b. Nineteen out of the forty horses were diagnosed by a veterinarian to have some levels of difficulty walking during the data collection. Those horses were diagnosed with either injury or inherent physical problems such as suspensory and soft tissue damage, sciatic issues, muscular lameness, and navicular syndrome.

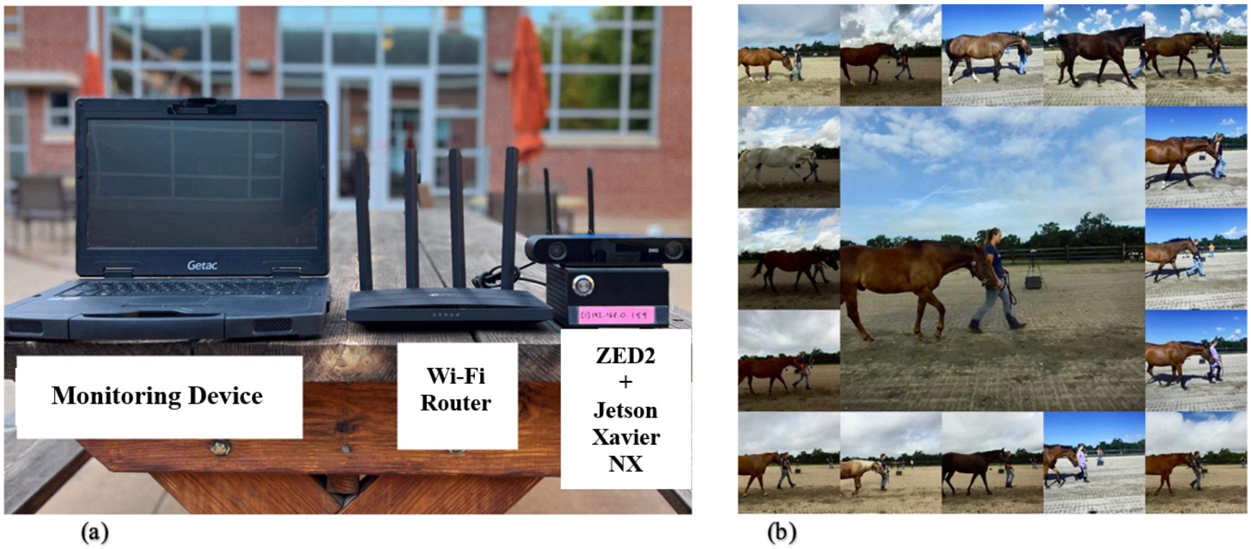


Figure 2-1. (a) Stereo video acquisition system: a ZED2 stereo camera, a Jetson Xavier NX embedded computer, a Wi-Fi router, and a monitoring device; and (b) sample images from the stereo video dataset that show different horse breeds under various lighting conditions.

Table 2-1 depicts the intrinsic parameters of the ZED2 stereo camera. Forty trotting or walking horses were video recorded at the Auburn University Equestrian Center (32.5856297220309, -85.5088532) during July of 2021.

Table 2-1. Intrinsic parameters of the used ZED2 stereo camera. Where C_x and C_y are the coordinates of the principal point of the left lens of the stereo camera in the image coordinate system. f is the focal length and T_x is the baseline between the two lenses of the ZED2 stereo camera.

Stereo Camera Parameters	**Values (Unit)**
f	1058.75 Pixels
C_x	1133.64 Pixels
C_y	659.75 Pixels
T_x	0.12 m

2.3.2 *Data Processing Pipeline*

The stereo video processing pipeline is composed of three main modules: body landmark detection, hoof stride phase detection, and stereo 3D reconstruction. As depicted in Figure 2-2, the first step is to detect the body landmarks from the 2D video frames using DLC. Next, each hoof and its stride phase (i.e., stance or swing) was detected using a Faster R-CNN (Ren et al., 2015) object detector. Lastly, the detected body landmarks were projected back to 3D space using a semi-global block matching algorithm (SGBM) (Hirschmuller, 2008). Two stride parameters, SL and SD, were used as case studies to evaluate the system. The implementation details are described in the following sections.

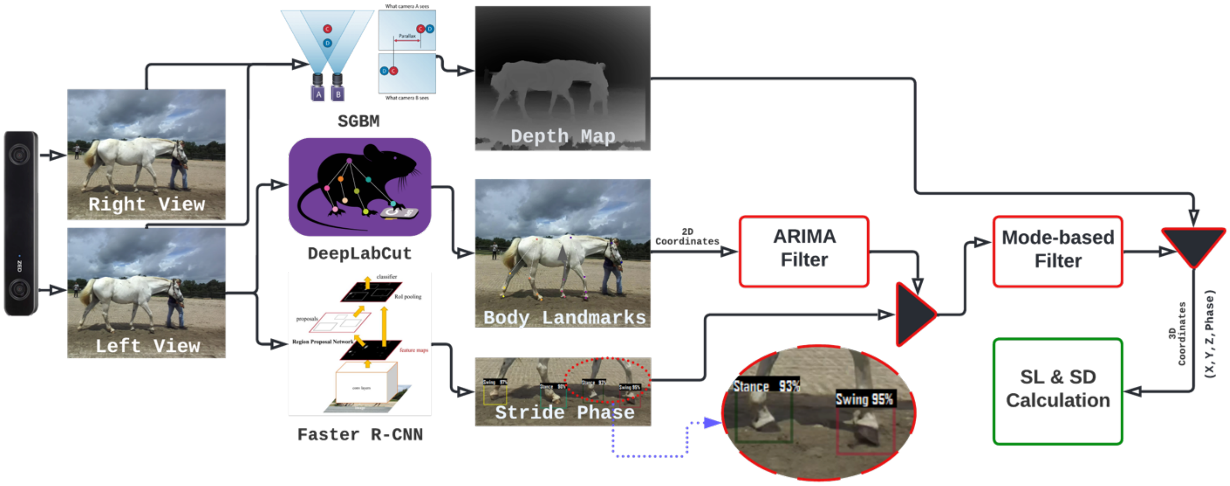


Figure 2-2. The proposed pipeline is designed to detect equine body landmarks, track those landmarks in the 3D space, and determine stride length and stance duration as case studies. Three points X, Y, and Z indicate how the 3D landmarks are positioned, specifically the hooves. In the gait cycle, Phase is the hoof phase, e.g., Stance or Swing.

2.3.3 *Body Landmark Detection and Filtering*

A DLC key point detection model was fine-tuned to detect equine body landmarks in individual video frames from the left lens of the stereo camera. Twenty frames were extracted from every video for annotation. DLC offers three methods to sample the frames: k-means clustering-based sampling, random sampling, and manual sampling. The k-means clustering method performs a k-means clustering of the pixel values in the video frames. A pre-defined number of frames are then sampled from the pool of grouped images. This sampling method aims to include different visual variations from all the available scenes. The k-means clustering method was used to provide equal opportunity for the different horse skin colors, backgrounds, and lighting conditions to be represented in the dataset. Such diverse dataset generates a model that is more robust to be applied in outdoor field conditions. The number of clusters was set to 10, and two frames were sampled from each cluster. The sixteen annotated body landmarks consisted of the hooves (4), the fetlocks (4), the knees (2), and the hocks (2), as well as nostril, poll, wither, and hip as shown in Table 2-2. For the DLC model 1200 frames were annotated, with 800 for model training and 400 for testing.

Table 2-2. Detailed descriptions of the DLC-based equine body landmark detection and Faster R-CNN-based gait phase detection datasets and models from 40 videos.

Model	Backbone	Pre-trained	Categories	Total	Tool	Training	Testing
DeepLabCut (DLC) Version 2.2	ResNet-101	ImageNet Dataset	Poll	810	DLC GUI Annotation Tool	20 Frames per Video, 800 Frames in Total	10 Frames per Video, 400 Frames in Total
			Withers	814			
			Nostril	803			
			Front Right Hoof	831			
			Front Right Fetlock	834			
			Front Right Knee	823			
			Front Left Hoof	823			
			Front Left Fetlock	842			
			Front Left Knee	841			
			Rear Right Hoof	834			
			Rear Right Fetlock	821			
			Rear Right Hock	845			
			Rear Left Hoof	812			
			Rear Left Fetlock	848			
			Rear Left Hock	813			
			Hip	810			
Faster R-CNN	ResNet-101 DC5 3X	COCO Dataset	Swing	1066	COCO Annotator	80% of Total	20% of Total
			Stance	2133			
			Occluded	24			

In many cases, landmarks such as withers, nostrils, and hips were obvious and relatively simple to annotate. However, landmarks such as the hoof, fetlock, hock, and knee of the left limbs could be occluded by the right limbs due to the side viewing angle. Body landmarks that were occluded were annotated based on the horse pose where they were most likely to be found. The key point detection performance was evaluated on seven landmarks (i.e., four hooves, nostril, poll, and wither) due to their paramount importance in equine gait analysis studies. ResNet101 was selected as the backbone for the DLC model. To train the model, the number of iterations was set to the value of 530,000 and a step-based learning rate decay method was used to set the learning rate to be 0.01, 0.005, and 0.002 for each one-third of the number of iterations. To add more variation to the training set, data augmentation was employed. A series of image transformations were applied to the annotated frames including gaussian noise (10%), elastic transformation, random rotation (0 to 15 deg), and motion blur. After an initial model training, the performance of the trained body landmark detector was evaluated using the test dataset. The detected landmarks were assessed both visually and based on the DLC-derived confidence level. If the detections were not associated with the correct body part or if they had a confidence level below 45%, they were selected to be annotated by a user to correctly update the body landmark. The newly annotated frames were then added to the previously developed training set. Finally, the pre-trained model was retrained with the new training dataset. The evaluation procedure described above was repeated three times to achieve a more accurate model.

As the DLC-predicted body landmark locations contain some errors and occasional extreme outliers, an autoregressive integrated moving average (ARIMA) filter in the post-processing tools of DLC was used to smooth the hoof trajectories as outlined and incorporated by (Mathis et al., 2018) in the DLC's framework. A nonseasonal ARIMA model is specified as $ARIMA(p, d, q)$, where p, d, and q refer to the number of autoregressive terms, the degrees of differentiations, and the number of lagged forecast errors, respectively (Kotu and Deshpande, 2019). An ARIMA (3, 0, 1) model was fitted to the landmark detection coordinates. Considering that the horses' motion was almost perpendicular to the camera, constant averaging with respect to one axis is appropriate. Plots of the sample autocorrelation function (ACF) and partial autocorrelation function (PACF) of the hoof trajectory were used to identify the orders. After the third lag, the PACF graph of the trajectory cut and that was used to determine the autoregressive term. The moving average (MA) term was determined by plotting the ACF. This filter was applied to generate a one-step-ahead forecast. The variance of the filtered trajectory was then calculated. For every landmark in the 2D space, the algorithm crossed out the ones that fall outside of the computed variance and replaced them with predicted coordinates.

2.3.4 *Hoof and Gait Phase Detection and Filtering*

To determine the gait phases of individual hooves, a pretrained Faster R-CNN model (Ren et al., 2015) with a Resnet-101 DC5 3X backbone in Detectron2 (Wu et al., 2019) was fine-tuned to detect the bounding boxes of hooves and the associated swing phase and stance phase of the gait. The number of epochs was set to 500 and the base learning rate was set to 0.001. The learning rate was decreased during training using a step-decay method with a minimum value of 0.0005. The size of each patch and the number of regions of interest per image were set to 2 and 128, respectively. COCO Annotator V0.11.1 (Brooks, 2019) was used to draw a bounding box around each hoof and assign a gait phase. If the sole of a hoof was visually determined to be more than five pixels away from the ground or to form an angle of five degrees or higher from the ground surface, that hoof was labeled as "Swing". Hooves that were in contact with the ground were labeled, as "Stance" as outlined by (Clayton, 2004). Hooves that were covered by another hoof were labeled as "Occluded". A series of data augmentation techniques (i.e., resizing, changing brightness, altering contrast and saturation, and flipping) were randomly applied to the frames to increase the generalization capacity of the trained model. As shown in Table 2, the numbers of annotated instances for "Swing", "Stance", and "Occluded" were 1066, 2133, and 24, respectively. For training the model, 80% of the total 810 frames were used and the remaining 20% was used to evaluate its performance.

2.3.5 *Gait Phase Assignment to Hoof Landmarks*

The Faster R-CNN-based gait phase detection results were assigned to the corresponding DLC-based hoof landmarks in each frame. Initially, all DLC-based hoof landmarks were labeled as "Neutral". The 2D coordinates were checked to see if the detected hooves were within the bounding box of the Faster R-CNN model output. For all the hooves, the label of the bounding boxes was associated with them, in which their DLC detected coordinates fell and had the shortest Euclidean distance with the center of the aforementioned bounding box. This procedure was applied to each hoof throughout the video. Using these two models together enabled the determination of the 2D coordinates of the hooves and their gait phase. This combined algorithm was evaluated by manually annotating the gait phases of each hoof in each frame of the 40-video dataset. The result was a comprehensive dataset with 6872 instances of "Stance", 4582 instances of "Swing", and 178 instances of "Occluded".

2.3.5.1 Hoof and Phase Detection Filtering

To measure distance in metric units (e.g., meter), a stereo 3D reconstruction of the 2D image landmark coordinates was needed. For each rectified stereo image pair, a disparity map was generated by using the semi-global block matching algorithm (SGBM) in the OpenCV library (Bradski, 2000). Note that stereo image rectification was automatically performed by the ZED software development kit (SDK) during image acquisition. SGBM has shown the ability to handle untextured areas with a smoothness term in its optimization objective function (Hirschmuller, 2008), which is suitable for horse bodies that have regions of homogenous coat colors. Considering the stereo camera baseline, its minimum imaging distance, a matching window size of 5 pixels, a disparity range of 100 pixels, the disparity map and the stereo camera parameters, the 2D landmark coordinates were projected back into 3D space using a set of equations (Equation 2-1):

$$\begin{cases} x = \dfrac{u - c_x}{W} \\ y = \dfrac{v - c_y}{W} \\ z = \dfrac{f}{W} \\ W = \dfrac{d}{T_x} \end{cases} \tag{2-1}$$

where x, y, and z are the coordinates of the landmark in 3D space, u and v are the coordinates in 2D image space of a landmark in the left image, W is a dummy variable and d is the calculated disparity for each pixel location. Optical center coordinates for the left lens ($\mathrm{c_x} = 1133.64$ pixels and $\mathrm{c_y} = 659.827$ pixels) were extracted from the ZED2 camera calibration file, as well as its focal length ($\mathrm{f} = 1058.75$ pixles) and baseline ($\mathrm{T_x} = 0.1200\mathrm{m}$).

2.3.6 *Stance Duration and Stride Length Estimation*

As a case study for the proposed pipeline, SL and SD were selected to be the output biomechanical parameters. SL and SD hold valuable information on the biomechanical soundness of horses as discussed previously. For instance, limb lameness has been shown to cause shorter stance duration and shorter stride length in horses (Barrey, 1999; Moorman et al., 2012; Nakamura et al., 2015; Peham et al., 2001; Serra Bragança et al., 2018). Literature has also shown that stride length correlates with horse size, so it is a reliable indirect measure of pipeline performance (Hole et al., 2002; Rooney et al., 1991). Thus, regression and pair-wise correlation were conducted to determine if an association exists between the pipeline's SL, SD and body measurements. SL and SD were examined in the horses diagnosed with lameness for outliers.

To determine the duration of a stance, if the same hoof phase was seen in more than two consecutive frames, those frames are counted as valid for that phase. In this way the number of frames in which each hoof was in the stance or swinging was determined. SD was calculated as the frame count for a stance phase divided by frame rate (i.e., 15 FPS). The average stance location was calculated in 3D space and a Euclidean distance between two consecutive stance locations was measured. Using this method, the SL of every detected hoof was calculated. An image level evaluation was conducted by manually annotating all the landmarks of interest, such as the hooves, nostrils, and poll, in all the consecutive frames in the 40-video dataset. It resulted in a dataset of 2915, 2907, 2901, and 2908 instances for "Right Front Hoof", "Left Front Hoof", "Right Rear Hoof", and "Left Front Hoof", respectively.

2.4 EVALUATION METHODS

Individual modules in the proposed data processing pipeline were assessed by using the metrics discussed below. Furthermore, ground truth values of SD and SL were obtained using manual annotations of landmarks and gait phases.

2.4.1 *Landmark Detection Assessment*

To assess the accuracy of the 2D keypoint detection algorithm (DLC) to correctly estimate the coordinates of the body landmarks, root mean square error (RMSE) and mean absolute error (MAE) were used. RMSE is calculated using by (Equation 2-2)

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{\left(\hat{Y}_i - Y_i\right)^2}{n}} \tag{2-2}$$

where $\hat{Y} = (\hat{u}, \hat{v})$ is the predicted 2D coordinates of the landmark in pixels, $Y = (u, v)$ is the manually labeled coordinates of the same landmark, n is the total number of the body landmark detections, and finally, i is the present frame in a sequence of frames. This RMSE value is uniformly averaged for u and v values to get one RMSE. Similarly, to calculate MAE, Equation 2-3 was utilized. The parameters are the same as Equation 2-2.

$$MAE = \frac{\left|\sum_{i=1}^{n}\left(\hat{Y}_i - Y_i\right)\right|}{n} \tag{2-3}$$

2.4.2 *Hoof and Phase Detection Evaluation*

Average precision (AP) (Zhu, 2004) was used to evaluate the accuracy of the Faster R-CNN-based hoof and gait phase detection model. In the field of object detection, AP is a popular metric for evaluating object detectors (e.g., Faster R-CNN and SSD) by calculating the average precision value over a range of recall values between 0 and 1. Here, AP is calculated as the area under the precision-recall curve, generated by Detectron2. The performance of Faster R-CNN in the hoof gait phase detection was assessed both in the training and testing datasets using the mean AP value for the "Swing" and "Stance" phases.

2.4.3 *Estimation of the Phase of a Hoof Using Coordinates Evaluation*

For each hoof coordinates that were identified by DLC, accuracy, precision, and recall of the associated hoof cycle phase were computed. These phases were determined by the trained Faster R-CNN model. To assess the efficiency of the post-processing procedures, these values were computed before and after the mode based and ARIMA filtering methods were applied. Equation 2-4 was used to calculate precision.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2-4}$$

where true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) were based upon manual labeling of the gait phases of hooves and comparing them to the outputs of the pipeline. Precision, recall, and F_1 were calculated using Equations 2-5, 2-6, and 2-7. Those metrics were used to assess the phase detection algorithm both before and after including the post-processing techniques.

$$Precision = \frac{TP}{Number\ of\ Prediction} \tag{2-5}$$

$$Recall = \frac{TP}{Number\ of\ Ground\ Truth\ Objects} \tag{2-6}$$

$$F_1 score = \frac{(Precision \times Recall)}{(Precision + Recall)} \tag{2-7}$$

2.4.4 *Stereo 3D Reconstruction Evaluation*

Head length was used to evaluate the stereo 3D reconstruction accuracy as it should not change throughout the videos. In-field measurement is fairly accurate compared to SL and SD which cannot be measured reliably with tape measures. In the field ground truth head length was measured from the two top corners of the nostrils straight to the front of the poll using a tape measure as defined in the previously mentioned protocol in data collection section. The accuracy of the 3D reconstruction module was assessed using a semi-automated method that used annotations from all the nostrils and polls body marks for all frames in the 40 videos collected for the dataset. A dataset of 2904 instances for nostril and 2912 instances for poll was created from the annotated images. In the semi-automated method, head length was computed as the Euclidean distance between nostril and poll in 3D space for each frame using manual image annotations and SGBM. Then, those measurements were averaged over all frames in a video for evaluation. The semi-automated head length aimed to assess the accuracy without considering the error introduced by DLC-based landmark detection. Similarly, fully automated head length estimation was also obtained using DLC results and SGBM. The two sets of computed values were then compared to in-field head length measurements. RMSE and MAE were used to assess the accuracy of the pipeline measurements against in-field data collection.

Additionally, Bland-Altman (B-A) analysis (Bland and Altman, 1999) was employed to evaluate bias and limits of agreement (LoA) between in-field manual measurements and semi- or fully automated head length estimations. In B-A, by generating a scatter plot of the differences between measurements of two systems is evaluated (Equation 2-8) as a function of the average of those measurements. Any systematic difference between the two measurement systems (i.e., bias), is quantified by the mean difference (Equation 2-9). Two LOAs lie at 1.96 standard deviations of difference above and below the bias with confidence intervals of 95% (Equations 2-10 to 2-12). The closer the bias is to zero, the more accurate the proposed method is against the reference method. If the LOAs are close to the bias line, the spread of differences is very small, suggesting that in most cases, the spread is very narrow. As part of our Bland-Altman plots, regression lines indicate whether bias was (mostly) constant over the measurements in our data set and whether homoscedasticity was present. A similar pattern of data presentation has been observed in other studies in this area (Bosch et al., 2018; Hatrisse et al., 2022).

$$Difference = Reference\ Method - \ Proposed\ Method \tag{2-8}$$

$$Bias = Mean\ (Difference) \tag{2-9}$$

$$SD = \ Standar\ Deviation(Difference) \tag{2-10}$$

$$LimitAgreementHigh\ (Upper\ LOA) = Bias + 1.96\ \times\ SD \tag{2-11}$$

$$LimitAgreementLow\ (Lower\ LOA) = Bias - 1.96 \times SD \tag{2-12}$$

The intra-class correlation (ICC) was also calculated to quantify the consistency between field and system-derived measurements (Barnhart et al., 2016). This step was done using the R package psychometric (version 2.2). The ICC yields values between zero and one. If ICC is zero, there is poor consistency between the two measurement systems, and if it is one, there is complete consistency. More specifically, ICC is calculated by (Equation 2-13)

$$ICC = \frac{t_{00}}{\left(t_{00} + \frac{\sigma^2}{n_j}\right)} \tag{2-13}$$

where t_{00} is the variance of the intercept of the model, σ^2 is the residual variance for the model, and n_j is the size of the population.

2.4.5 *Stride Length and Stance Duration Evaluation*

SL and SD estimations were evaluated by annotating the 2D coordinates of the hooves in all frames of the videos and finding their 3D coordinates via SGBM. A manually annotated dataset of hooves' phases was used, as described in the "Gait phase assignment to hoof landmarks" section, to calculate the SD and consecutively SL of the landmarks. A comparison was then conducted between the ground truth measurements and the output of the pipeline through B-A analysis. An ICC analysis was also conducted on the measured values for SL and SD. Revisions were also made to equations 2-2 and 2-3 in order to make them 3D. In 3D space, $\hat{Y} = (\hat{X}, \hat{Y}, \hat{Z})$ represents the predicted coordinates of a landmark. As well as $Y = (X, Y, Z)$ being the manually annotated landmarks that were projected into 3D space.

2.4.6 *Experiment Environment*

Model training and data analyses were performed on a workstation equipped with an AMD Ryzen Threadripper 2970WX 4.2 GHz 24-Core Processor, 64 GB RAM, and two Nvidia Titan RTX GPUs with 48 GB of VRAM, running a Linux Ubuntu 20.04.4 LTS operating system.

2.5 RESULTS

2.5.1 *Landmark Detection*

The training loss of the DLC landmark detection model was 0.002 pixels. For the testing dataset, the RMSE and MAE for landmark detection without an ARIMA filter were 8.24 pixels and 6.52 pixels. Using the ARIMA filter, the RMSE and MAE were reduced to 5.14 pixels and 4.87 pixels. The detection errors of DLC equipped with an ARIMA filter of the seven individual landmarks for both the training and the testing datasets are shown in Figure 2-3. Higher means of detection error can be seen for body landmarks that were periodically occluded (i.e., left hooves). For instance, the left rear hoof had an error of 16 pixels as compared to 10 pixels for the right rear hoof. Nostril detection had the lowest mean error of 3.00 pixels in the training dataset and 5.00 pixels in the testing set. As qualitative results, Figure 2-4 depicts examples of the DLC detected landmarks in the testing dataset. The 2D coordinates of the landmarks were determined and marked by different colors.

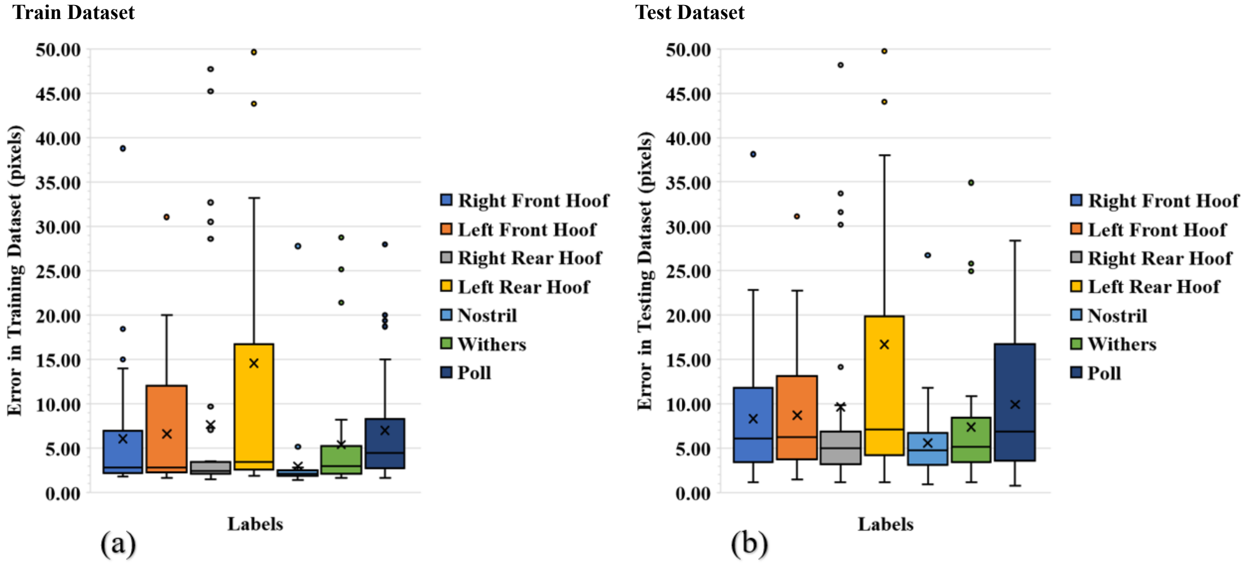


Figure 2-3. DLC detection errors for the (a) training dataset and (b) testing dataset for seven horse body landmarks. The cross mark in each boxplot stands in for the mean, the horizontal center line for the medians, and the outside-box dots for the outliers. Each boxplot's top and lower boundaries correspond to the 25% and 75% percentiles.

Figure 2-4. A visual evaluation of the performance of DLC in detecting body landmarks; (a) shows how DLC detects landmarks on the horse's body thoroughly. The hoof and fetlock on the rear left side of (b) are occluded and therefore not picked up by DLC. Likewise, the knee and hock on the left side of (c) are similarly not detected.

2.5.2 *Hoof and Gait Phase Detection*

Training the Faster R-CNN model was done with the specified dataset discussed in Table 2-2, and Figure 2-5 depicts the training losses. Figure 2-5 depicts significant losses in the hoof phase detection algorithm, i.e., classification and box regression losses during the training session. The value of classification accuracy increased during training, starting from around 0.85 and approaching 1.0. The performance of the Faster R-CNN model was evaluated with the test dataset and yielded an AP of 33.196 and 39.714 for the swing phase and stance phase, respectively. A visual representation of the phase detections algorithm applied on an instance from the testing dataset, in conjunction with their corresponding bounding boxes, is shown in Figure 2-6.

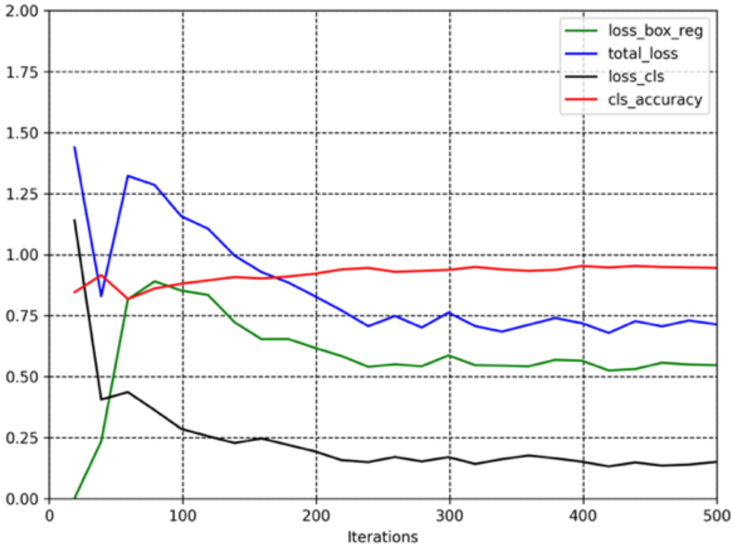


Figure 2-5. Total loss, class accuracy, box regression loss, and classification loss of Faster R-CNN model for hoof and gait phase detection.

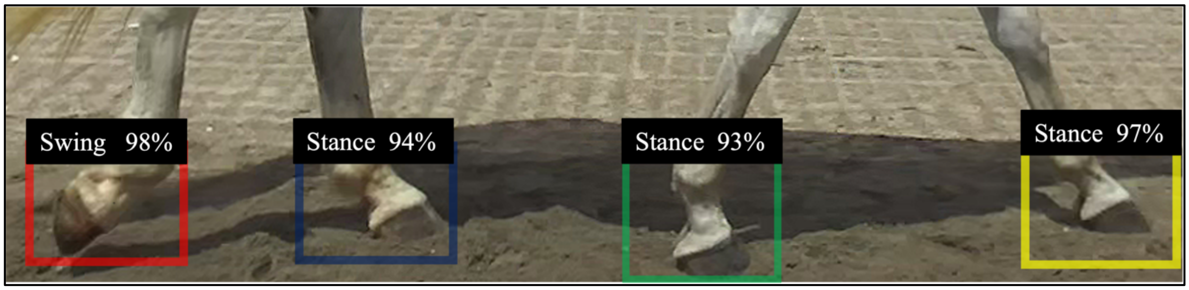


Figure 2-6. Bounding boxes around individual hooves and gait phase classification generated by the trained Faster R-CNN model.

2.5.3 *Estimating the Phase of a Hoof Using Coordinates*

The following results were computed using different combinations of the proposed modules in the previous sections. As shown in Table 2-3, the precision increased from 0.56 using DLC and Faster R-CNN models alone to 0.83 by applying the additional ARIMA and median filters. Recall and F1 scores had similar improvements, increasing from 0.58 and 0.55 to 0.95 and 0.78, respectively.

Table 2-3. Precision, recall, and F1 score for different gait phase classification methods.

Gait Phase Classification Method	**Precision**	**Recall**	**F1 Score**
DLC + Faster R-CNN	0.56	0.58	0.56
DLC w/ARIMA Filter + Faster R-CNN	0.61	0.74	0.61
DLC w/ ARIMA Filter + Faster R-CNN w/ Median Filter	0.83	0.95	0.78

The hoof gait phase detection algorithm had a significant improvement when the filtering algorithms were incorporated. Figure 2-7 illustrates the confusion matrices for gait phase detection using different combinations of trajectory generation algorithms and filters. Both in "Swing" and "Stance", true positive rates were increased from 56.22% to 61.20% in "Stance" and from 18.92% to 30.95% in "Swing" detections.

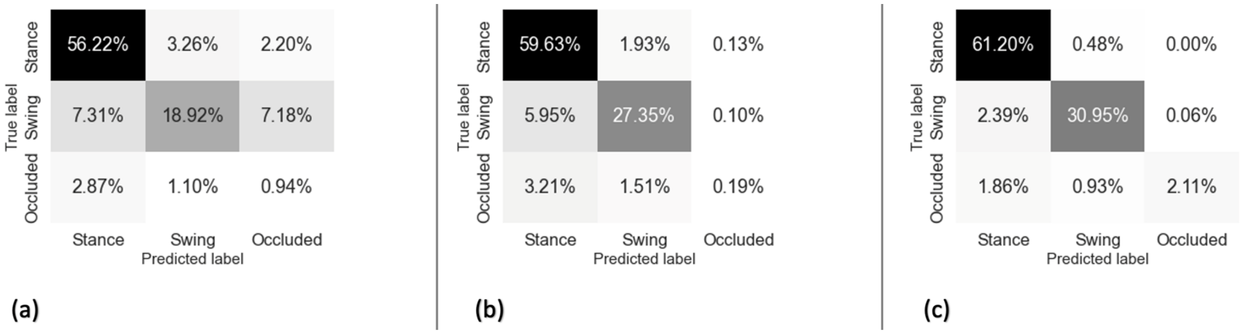


Figure 2-7. Confusion matrices of hoof gait phase prediction using different combinations of algorithms: (a) DLC + Faster R-CNN; (b) DLC w/ ARIMA filter + Faster R-CNN; (c) DLC w/ ARIMA filter + Faster R-CNN w/ mode-based filter.

2.5.4 *3D Reconstruction and Measurements of Head Length*

The RMSE and MAE for head length estimation using manually annotated nostril and poll landmarks along with the SGBM algorithm were 0.019 m and 0.017 m, respectively. When the manual annotations were changed to the DLC predictions with ARIMA filtering, the system performance decreased with an RMSE and an MAE being 0.028 m and 0.022 m, respectively. B-

A plots for semi- and fully automated head length estimation methods against the ground truth method were shown in Figure 2-8. Table 2-4 summarizes the biases, LOAs, and ICCs for the two evaluated methods. The fully automated method resulted in a larger standard deviation than the semi-automated method. On the other hand, the biases of both methods were similar. As for ICC, the semi-automated method was excellently consistent with the ground truth method (ICC = 0.95), whereas the fully automated method was moderately consistent (ICC = 0.72) according to the guideline by (Koo and Li, 2016). P-values of the slope of the fitted regression lines in Figure 2-8 (a) semi-automated method and (b) fully automated were 0.939 and 0.059, respectively. Because both p-values were greater than 0.05, proportional bias is not statistically significant in the two B-A plots and thus, the regression lines were not shown.

Table 2-4. Summary of head length measurement agreement statistics (B-A and ICC) when compared to in-field measurements. ICC: intra-class correlation; LoA: limit of agreement. The semi-automated method uses manual annotations in conjunction with SGBM, while the fully automated method uses SGBM in conjunction with DLC's output.

	Semi-automated	**Fully automated**
Upper LoA (m)	0.0020	0.0345
Lower LoA (m)	-0.0360	-0.0641
Bias (m)	-0.0169	-0.0147
ICC	0.95	0.72

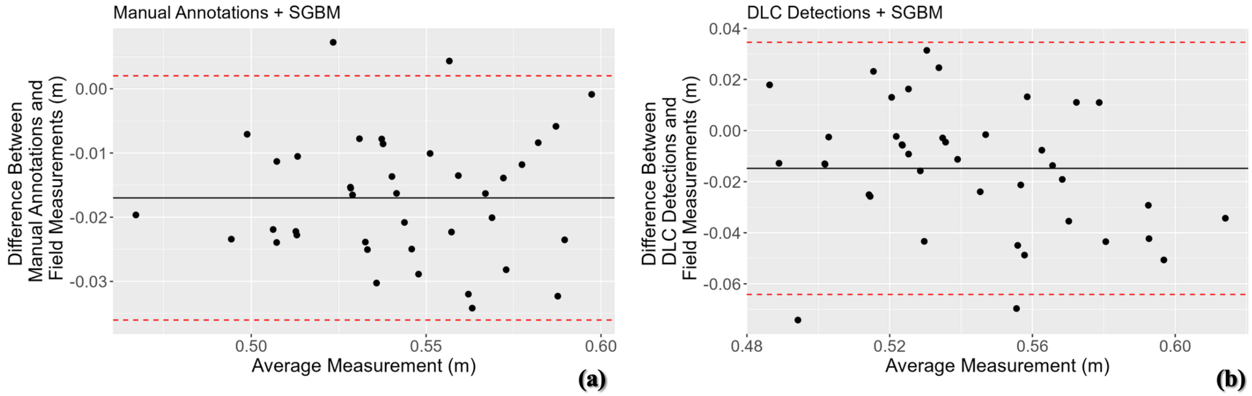


Figure 2-8. Bland-Altman analysis on estimated head length measurements against ground truth manual measurements; (a) manual annotation with SGBM, and (b) detected landmarks using DLC with SGBM.

2.5.5 *Stride Length Stance Duration Evaluation*

The system-derived SD resulted in an RMSE of 0.094 s in comparison to the manual measurements from the videos. In 2D image space, RMSE between the system-derived SL and the ground truth was 5.73 pixels. The B-A analysis shows a bias of -0.025 s for SD (Figure 2-9a). The SL B-A graph (Figure 2-9b) had a bias and a standard deviation of -0.042 m and 0.041 m, respectively. SD's upper and lower LOAs were 0.191 s and -0.241 s, respectively. Also, the upper and lower LOAs of SL were 0.040 m and -0.124 m. According to Figure 2-9a, the slope of the fitted regression line had a p-value of 3.853×10^{-6} for SD graph. The regression line does indicate a clear trend and after visual assessment, there is evidence of heteroscedasticity. On the other hand, in Figure 9 (b) for SL, the P-value of the slope of the fitted regression line was 0.071, and thus not shown in the figure. Table 2-5 summarizes the results of B-A analyses and ICCs for SD and SL estimation. ICC for total stance durations was 0.79, while for stride length it was 0.98, showing a moderately consistent behavior in measuring SD and an excellent one in computing SL.

Table 2-5. Summary of SL and SD measurement agreement statistics (B-A and ICC) when compared to manual image-level annotations measurements. ICC: intra-class correlation; LoA: limit of agreement; RF: right front; LF: left front; RR: rear right; LR: left rear.

	Stance Duration	**Stride Length**
Upper LoA	0.191s	0.040m
Lower LoA	-0.241s	-0.124m
Bias	-0.025s	-0.042m
ICC RF	0.86	0.99
ICC LF	0.73	0.98
ICC RR	0.86	0.98
ICC LR	0.72	0.98
ICC Total	0.79	0.98

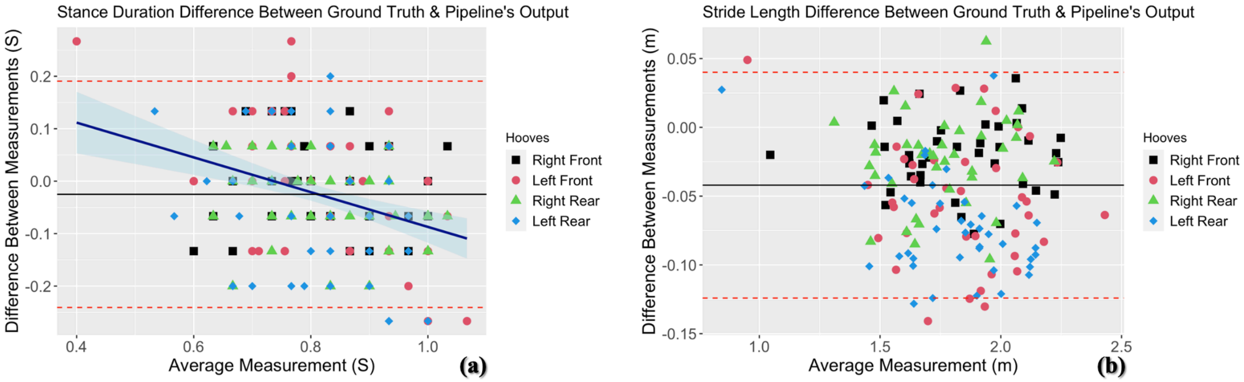


Figure 2-9. Bland-Altman analyses for (a) stance time (SD) and (b) stride length (SL) estimations.

There were 7 horses with stance durations that differed between legs among the 19 horses diagnosed as lame. As an example, three horses affected by injuries to the suspensory ligaments, deep digital flexor tendons, and tendon sheaths are shown in Figure 2-10 with their stance durations. Among the other 12 horses with some degree of difficulty walking, there were no noticeable differences in stance duration.

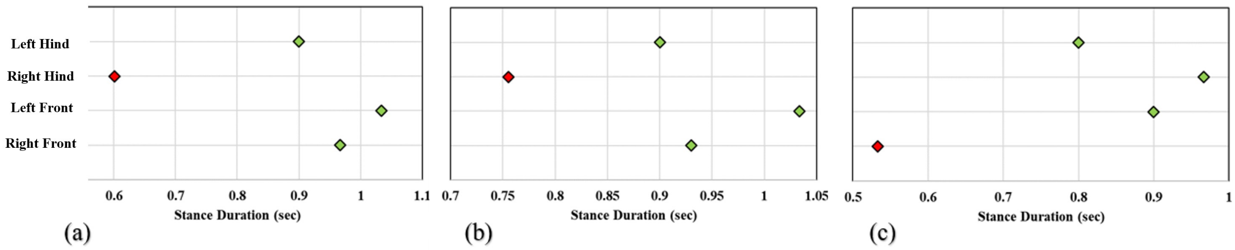


Figure 2-10. The average stance duration in three subjects with lameness in one of the legs based on a veterinarian's diagnosis: (a) injury to the deep digital flexor tendon in a right hind leg, (b) suspensory injury to the right hind leg, and (c) issues with a right front tendon sheath.

SL had a positive correlation with pastern length ($R^2 = 0.53$), demonstrating greater stride length with longer pastern length (Figure 2-11).

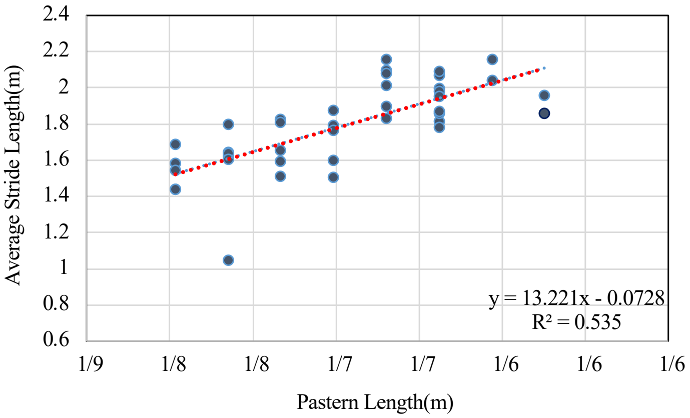


Figure 2-11. Average stride length (m) and pastern length (m) of the forty horses in the dataset with a fitted linear regression model to the data points.

2.6 DISCUSSION

2.6.1 *Performance of the Pipeline*

This study demonstrated that accurate gait classification using Faster R-CNN in conjunction with marker-less body landmark detection and stereo videography to determine different biomechanical properties of horses can be achieved using a high throughput and cost-effective setup and pipeline. Traditionally, the human eye has served as a common method assessing gait. However, human subjective assessment proved suboptimal because it could be affected by the temporal limitations of the human eye i.e., having a limited frequency response. With the proposed system, researchers will be enabled to achieve quantitative results on different biomechanical parameters such as SL and SD under field conditions.

It can be observed that DLC eventually fails to detect hoofs in a single equine locomotion video. There are many reasons why this can happen, including when another limb, an object on the track, or flying dirt from the horse's foot impacting the ground obscures a hoof. Figure 2-5 shows that some of the body landmarks in the middle image, such as the left rear hoof, was occluded by the right body part, and thus, the DLC failed to detect it. A further problem is the detection of two hooves when they overlap (e.g., two front hooves). Failures like these usually last for a relatively small number of frames compared to the total number of frames taken during a stride. By incorporating an ARIMA filter, the RMSE of the landmark detection algorithm was reduced from 8.24 pixels to 5.14 pixels. Potentially this could be caused by removing the outlier detections and replacing them with more reasonable values. Both the training and testing datasets had similar mean detection errors for different body landmarks. Meanwhile, right-sided landmarks were detected more accurately than left-sided landmarks as shown in Figure 2-4. This is likely caused by periodic occlusions from the right-side imaging angle that affected the left-side hooves. Capturing and processing a second video from the left side of the horse can improve detection accuracy for the left hooves. Out of the seven landmarks evaluated, detection of the poll had the second-largest error range. There was high variability in the poll appearance due to head movement in the video frames. There is a possibility of alleviating this issue by adding more annotations to the training dataset.

It was found that the proposed Faster R-CNN model for hoof and gait phase detection had an average AP value of 36.455. Employing a trained Faster R-CNN model to do the fine-grain classification of the hoof phase might be beneficial in two ways. Using this method, researchers get more intuitive results by observing the gait classes directly rather than relying on indirect measurements like the trajectory of a hoof to calculate the gait stages (Clayton, 2004). It is also important to note that extracting phases just by assessing IMU or other physical sensors may be extremely dependent on the location of the sensors on the horses, which may differ from study to study (Moorman et al., 2012).

However, due to occluded hooves in some frames, it was difficult to determine if the hoof was swinging. In this case, if only DLC coordinates coupled with Faster R-CNN bounding boxes without any post-processing algorithms were used, precision = 0.56 and recall = 0.58. Though, when a median-based filter on hoof phase detections and an ARIMA filter for DLC detections were applied, the accuracy and recall of the final detections were 0.83 and 0.95, which indicates the improvement in performance due to the post-processing algorithm.

Some methods available in the literature tried to suggest an alternative system to OMC such as EquiMoves which incorporates inertial sensors, or GAITRite mat with embedded pressure sensors. (Bosch et al., 2018; Cutlip et al., 2000) demonstrated these systems have within mm level accuracy when evaluated by the results of OMC. Although these techniques offer accurate readings, they are not flexible enough to modify the number of landmarks of interest after data collection or to accommodate various data collection settings. Also, in physical sensor-based systems such as IMUs, the final measurements are significantly influenced by the sensor placement on the body (Moorman et al., 2012). In comparison, our proposed pipeline is not limited by fixed markers, arduous placement, and inconsistent results if the markers are not accurately positioned; these advantages were explained by (Mathis et al., 2018b; T. Nakamura et al., 2016; Nath et al., 2019) when deep learning methods are incorporated. This system could also be potentially easier for researchers to use under field conditions and serve as a high-throughput and cost-effective solution. Moreover, 2D videography has limitations when it comes to measuring kinematic gait parameters. To create metric unit measurements in 2D techniques, the observed gait parameters must be normalized by a known length in the collected photos, which might be error-prone (Vinika Gupta, 2021). Additionally, a stereo 3D camera does not need to be positioned perpendicular to the horse's movement direction, giving end users more flexibility. Our suggested pipeline, similar to physical sensor-based techniques that incorporate physical devices, delivers landmark trajectories in the 3D space and by metric units. Furthermore, the advantage of using a commercial stereo camera (ZED2) whose views have already been rectified and calibrated means that the developed method does not require time-consuming calibration procedures, as opposed to using multiple camera systems in markerless 3D pose estimation methods like AniPose and FreiPose (Karashchuk, Rupp, et al., 2021; Zimmermann et al., 2020).

This work's logistical inability to use an optical measurement system (the gold standard) to confirm the precision of the proposed pipeline for landmark identification and stride length estimation is one of its limitations. An optical measurement system was not used because the actual reflecting indicators would conceal the body landmarks and also raised the risk of disclosing actual locations to DLC, therefore the model might learn the position of the markers rather than real body landmarks. If the two procedures were employed sequentially, placing, and removing the markers might potentially change the horse's behavior and, as a result, the final evaluations might not be relevant. Given this constraint, head length was measured in this study to indirectly assess the ability of the system in 3D measurements and consecutively, stride length estimation. This procedure was carefully crafted and resulted in somewhat less accurate findings when DLC's outputs were used in the fully automated approach as compared to the semi-automated method. One probable reason for this behavior is DLC's decreased accuracy as compared to manual annotations in establishing the exact location of body landmarks, as well as occasional misdetections of the landmarks. As seen in Figure 2-8, DLC combined with SGBM had a bias of 0.014 m and a standard deviation of 0.025 m which still could be considered accurate as the range of the head length was between 0.45 m and 0.63 m. This step may well illustrate the pipeline's potential performance when further 3D kinematic parameter measurements are made. Figure 2-8 also shows that, while the biases in manually annotated images and DLC outputted images are similar, the standard deviation for DLC results are greater. This may be attributed to the erratic behavior of detections in the landmark detector model. For both manually labeled images and DLC derived ones, the regression lines in that figure reveal no distinct trend as the p-value of their slopes were greater than 0.05 with no visible sign of heteroscedasticity. The greater ICC for the approach with manual annotations than the DLC-based method can be explained by the hectic behavior of DLC when detecting the landmarks of interest.

Furthermore, when compared to manual frame annotations for stance duration, an R^2 of 0.93 and an RMSE of 0.094 s were obtained which indicates the system's sensitivity to variations in the gait cycle. Running a B-A analysis on the final products of the pipeline (illustrated in Figure 2-9) shows that only SL with a p-value of below 0.05 had a negative slope for mean errors. This behavior may be explained by the trained Faster R-CNN model's miss detections or occasional miss classification of the phases of the hooves when they are occluded. Additionally, the combination of the pipeline modules may be to blame for the underestimate in 3D, particularly for the body landmarks on the left that are repeatedly covered in the videos. When compared to human assessments, ICC values for SL and SD demonstrated notable consistency in measurements. Table 2-5 shows that using the method to detect right-side landmarks could result in more consistent results, possibly due to fewer occlusions and better performance of both models that are responsible for detection of landmarks and classification of hoof phases.

This system can be exploited further to detect musculoskeletal diseases (e.g., lameness). It has been demonstrated that a model-based behavioral analysis can be run on horse motion to extract and detect lameness in equestrian animals (Li et al., 2021). Additionally, it might assist in breeding programs by providing quantitative data on biomechanical traits to select progeny of interest.

Our evaluation of the outputs of the pipeline will provide insight for future research into whether different SDs are detected in horses with lameness. Among the 19 lame horses sampled, there were seven horses with different stance durations between the hooves, indicating there is potential for this system to be used in lameness detection. In some instances, the kind of lameness a horse exhibited in our dataset was not correlated with the length of stance in the study. The type of lameness is vital in the way the pipeline needs to be exploited and will be incorporated in future work. Likewise, the average system-derived stride length in 3D space was found to have an R^2 of 0.55 with the manually measured pastern length as illustrated in Figure 2-11, showing a consistent result with the available studies in the literature (Baban et al., 2009; Heglund and Taylor, 1988; Heglund et al., 1974; Sánchez et al., 2013). However, the scope of this work did not include lameness detection; the goal was to develop and evaluate a measurement system.

The stereo imaging pipeline can be extended to multiple wirelessly synchronized stereo cameras. This enables tracking and analysis of landmarks around a horse's body. Also, a more precise stereo matching algorithm can be utilized to improve the accuracy of the stereo matching and the quality of disparity maps. SGBM was the preferred method due to its high efficiency considering the computational system that was used. However, state-of-the-art CNNs (Chen and Jung, 2018) can be used to further improve the processes of computing the similarity between stereo image patches. Furthermore, since the landmarks' 3D coordinates are what is important in this process and their disparity might perhaps be determined directly, dense stereo matching may not be required. Besides, the distance of the camera to the horse determines the spatial resolution (i.e., detail level) and the accuracy of stereo matching-based depth estimation, therefore another area of interest is to study how camera-horse distance impacts the performance of the proposed system.

2.7 CONCLUSION

An automated EKGA technology was developed that consists of 3D stereo videography system in conjunction with a processing pipeline. This EKGA system was built by combining several modules to identify body landmarks (DLC model), to classify of hoof cycle phases (Faster R-CNN model), and ultimately to compute 3D coordinates of the detected landmarks (SGBM). SL and SD were selected as the pipeline's outputs as a case study for the proposed methodology, and post-processing techniques were applied to boost the performance of the modules. According to in-field head length measurements, the 2D detection algorithm performed satisfactorily and consistently when combined with the stereo matching algorithm. The fine grain hooves' classification was performed precisely in this pipeline. Following a B-A study, it is evident that the suggested combination of modules could potentially be used in evaluating kinematic parameters of horses; with a bias of 0.11 s for SD and -0.014 m for SL.

Research in gait genetics could benefit from our developed pipeline used in this study. Equine genetic studies focusing on locomotion, such as gait or sports performance, require precise phenotyping to identify trends in populations or subpopulations. Overall, we think that the proposed method can be developed into a low-cost, practical, and rapid analytical tool for animal scientists to perform EKGA. Even though this system sometimes might be less precise than marker-based optical motion capture systems available in the market, it requires less efforts to use. In most of the available studies, gait phenotyping is still done subjectively, making it less accurate than desirable. Therefore, we believe that our more accurate methods will make it possible to learn more about how genetics shapes gait.

In order to improve the accuracy of the kinematic gait parameters, notably SL and SD measurements, additional assessment algorithms and data processing methods, such as a combination of multiple cameras, must be investigated.

2.8 ACKNOWLEDGMENTS

This project was funded by the Auburn University Intramural Grants Program and the College of Agriculture.

2.9 REFERENCES

Arkell, M., Archer, R. M., Guitian, F. J., May, S. A. (2006). Evidence of bias affecting the interpretation of the results of local anaesthetic nerve blocks when assessing lameness in horses. Veterinary Record, 159(11), 346–348. https://doi.org/10.1136/VR.159.11.346

Baban, M., Curik, I., Antunovic, B., Cacic, M., Korabi, N., Mijic, P. (2009). Phenotypic Correlations of Stride Traits and Body Measurements in Lipizzaner Stallions and Mares. Journal of Equine Veterinary Science, 29(6), 513–518. https://doi.org/10.1016/J.JEVS.2009.04.193

Bala, P. C., Eisenreich, B. R., Bum, S., Yoo, M., Hayden, B. Y., Park, H. S., Zimmermann, J., and Hayden, B. (2020). OpenMonkeyStudio: Automated Markerless Pose Estimation in Freely Moving Macaques. BioRxiv, 2020.01.31.928861. https://doi.org/10.1101/2020.01.31.928861

Barnhart, H. X., Yow, E., Crowley, A. L., Daubert, M. A., Rabineau, D., Bigelow, R., Pencina, M., and Douglas, P. S. (2016). Choice of agreement indices for assessing and improving measurement reproducibility in a core laboratory setting. Statistical Methods in Medical Research, 25(6), 2939–2958. https://doi.org/10.1177/0962280214534651

Barrey, E. (1999). Methods, Applications and Limitations of Gait Analysis in Horses. The Veterinary Journal, 157(1), 7–22. https://doi.org/10.1053/TVJL.1998.0297

Barrey, E., Auvinet, B., Couroucé, A. (1995). Gait evaluation of race trotters using an accelerometric device. Equine Veterinary Journal, 27(S18), 156–160. https://doi.org/10.1111/j.2042-3306.1995.tb04910.x

Behnke, R., Nakirya, M. (2012). The contribution of livestock to the Ugandan economy. IGAD Centre for Pastoral Areas and Livestock Development, (02).

Bland, J. M., Altman, D. G. (1999). Measuring agreement in method comparison studies. Statistical Methods in Medical Research, 8(2), 135–160. https://doi.org/10.1177/096228029900800204

Bosch, S., Serra Bragança, F., Marin-Perianu, M., Marin-Perianu, R., van der Zwaag, B., Voskamp, J., Back, W., van Weeren, R., and Havinga, P. (2018). EquiMoves: A Wireless Networked Inertial Measurement System for Objective Examination of Horse Gait. Sensors, 18(3), 850. https://doi.org/10.3390/s18030850

Bradski, G. (2000). The OpenCV Library. Dr. Dobb's Journal of Software Tools. https://opencv.org/

Brooks, J. (2019). COCO Annotator. (Version 0.11.1). https://github.com/jsbroks/coco-annotator

Brooks, S. A., Makvandi-Nejad, S., Chu, E., Allen, J. J., Streeter, C., Gu, E., McCleery, B., Murphy, B. A., Bellone, R., and Sutter, N. B. (2010). Morphological variation in the horse: defining complex traits of body size and shape. Animal Genetics, 41(SUPPL. 2), 159–165. https://doi.org/10.1111/J.1365-2052.2010.02127.X

Chen, B., Jung, C. (2018). Patch-Based Stereo Matching Using 3D Convolutional Neural

Networks. Proceedings - International Conference on Image Processing, ICIP, 3633–3637. https://doi.org/10.1109/ICIP.2018.8451527

Clayton, H. (2004). The dynamic horse: A biomechanical guide to equine movement and performance. Sport Horse Pub. Cutlip, R. G., Mancinelli, C., Huber, F., Dipasquale, J. (2000). Evaluation of an instrumented walkway for measurement of the kinematic parameters of gait. Gait and Posture, 12(2), 134–138. https://doi.org/10.1016/S0966-6362(00)00062-X

Dunn, T. W., Marshall, J. D., Severson, K. S., Aldarondo, D. E., Hildebrand, D. G. C., Chettih, S. N., Wang, W. L., Gellis, A. J., Carlson, D. E., Aronov, D., Freiwald, W. A., Wang, F., and Ölveczky, B. P. (2021). Geometric deep learning enables 3D kinematic profiling across species and environments. Nature Methods, 18(5), 564–573. https://doi.org/10.1038/S41592-021-01106-6

Grice, A., and others (2018). 2017 American Horse Council economic impact study. In Proceedings of the 64th Annual Convention of the American Association of Equine Practitioners, San Francisco, California, USA, 1-5 December 2018 (pp. 502–504).

Günel, S., Rhodin, H., Morales, D., Campagnolo, J., Ramdya, P., Fua, P. (2019). Deepfly3D, a deep learning-based approach for 3D limb and appendage tracking in tethered, adult Drosophila. ELife, 8. https://doi.org/10.7554/ELIFE.48571

Hagen, J., Jung, F. T., Brouwer, J., Bos, R. (2021). Detection of Equine Hoof Motion by Using a Hoof-Mounted Inertial Measurement Unit Sensor in Comparison to Examinations with an Optoelectronic Technique - A Pilot Study. Journal of Equine Veterinary Science, 101. https://doi.org/10.1016/j.jevs.2021.103454

Hatrisse, C., Macaire, C., Sapone, M., Hebert, C., Hanne-Poujade, S., de Azevedo, E., Marin, F., Martin, P., and Chateau, H. (2022). Stance Phase Detection by Inertial Measurement Unit

Placed on the Metacarpus of Horses Trotting on Hard and Soft Straight Lines and Circles. Sensors, 22(3), 703. https://doi.org/10.3390/S22030703/S1

Heglund, N. C., Taylor, C. R. (1988). Speed, stride frequency and energy cost per stride: how do they change with body size and gait? Journal of Experimental Biology, 138(1), 301–318. https://doi.org/10.1242/JEB.138.1.301

Heglund, Norman C., Taylor, C. R., Mcmahon, T. A. (1974). Scaling Stride Frequency and Gait to Animal Size: Mice to Horses. Science, 186(4169), 1112–1113. https://doi.org/10.1126/SCIENCE.186.4169.1112

Hirschmuller, H. (2008). Stereo Processing by Semiglobal Matching and Mutual Information. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(2), 328–341. https://doi.org/10.1109/TPAMI.2007.1166

Hole, S. L., Clayton, H. M., Lanovaz, J. L. (2002). A note on the linear and temporal stride kinematics of Olympic show jumping horses between two fences. Applied Animal Behaviour Science, 75(4), 317–323. https://doi.org/10.1016/S0168-1591(01)00194-0

José Sánchez, M., Dolores Gómez, M., Peña, F., García Monterde, J., Luís Morales, J., Molina, A., Valera, M. (2013). Relationship between conformation traits and gait characteristics in Pura Raza Español horses. Archives Animal Breeding, 56(1), 137–148. https://doi.org/10.7482/0003-9438-56-013

Karashchuk, P., Rupp, K. L., Dickinson, E. S., Azim, E., Brunton, B. W., Tuthill, J. C., Walling-Bell, S., and Sanders, E. (2021). Anipose: A toolkit for robust markerless 3D pose estimation. Cell Reports, 36. https://doi.org/10.1016/j.celrep.2021.109730

Karashchuk, P., Tuthill, J. C., Brunton, B. W. (2021). The DANNCE of the rats: a new toolkit for 3D tracking of animal behavior. Nature Methods, 18(5), 460–462.

https://doi.org/10.1038/s41592-021-01110-w

Keegan, K. G. (2007). Evidence-Based Lameness Detection and Quantification. Veterinary Clinics of North America: Equine Practice, 23(2), 403–423. https://doi.org/10.1016/J.CVEQ.2007.04.008

Keegan, K. G., Wilson, D. A., Smith, B. K., Wilson, D. J. (2000). Changes in kinematic variables observed during pressure-induced forelimb lameness in adult horses trotting on a treadmill. American Journal of Veterinary Research, 61(6), 612–619. https://doi.org/10.2460/ajvr.2000.61.612

Koo, T. K., Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. Journal of Chiropractic Medicine, 15(2), 155. https://doi.org/10.1016/J.JCM.2016.02.012

Kotu, V., Deshpande, B. (2019). Time Series Forecasting. Data Science, 395–445. https://doi.org/10.1016/B978-0-12-814761-0.00012-5

Li, C., Ghorbani, N., Broomé, S., Rashid, M., Black, M. J., Hernlund, E., … Cil, S. (2021). hSMAL: Detailed Horse Shape and Pose Reconstruction for Motion Pattern Recognition. https://doi.org/10.48550/arxiv.2106.10102

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nature Neuroscience, 21(9), 1281–1289. https://doi.org/10.1038/s41593-018-0209-y

Moorman, V. J., Reiser II, R. F., McIlwraith, C. W., Kawcak, C. E. (2012). Validation of an equine inertial measurement unit system in clinically normal horses during walking and trotting. American Journal of Veterinary Research, 73(8), 1160–1170.

https://doi.org/10.2460/ajvr.73.8.1160

Nakamura, A., Funaya, H., Uezono, N., Nakashima, K., Ishida, Y., Suzuki, T., Wakana, S., and Shibata, T. (2015). Low-cost three-dimensional gait analysis system for mice with an infrared depth sensor. Neuroscience Research, 100, 55–62. https://doi.org/10.1016/j.neures.2015.06.006

Nakamura, T., Matsumoto, J., Nishimaru, H., Bretas, R. V., Takamura, Y., Hori, E., Ono, T., and Nishijo, H. (2016). A Markerless 3D Computerized Motion Capture System Incorporating a Skeleton Model for Monkeys. PLOS ONE, 11(11), e0166154. https://doi.org/10.1371/journal.pone.0166154

Nath, T., Mathis, A., Chen, A. C., Patel, A., Bethge, M., Mathis, M. W. (2019). Using DeepLabCut for 3D markerless pose estimation across species and behaviors. Nature Protocols, 14(7), 2152–2176. https://doi.org/10.1038/s41596-019-0176-0

Peham, C., Licka, T., Girtler, D., Scheidl, M. (2001). The Influence of Lameness on Equine Stride Length Consistency*. The Veterinary Journal, 162(2), 153–157. https://doi.org/10.1053/tvjl.2001.0593

Pereira, T. D., Aldarondo, D. E., Willmore, L., Kislin, M., Wang, S. S.-H., Murthy, M., Shaevitz, J. W. (2019). Fast animal pose estimation using deep neural networks. Nature Methods, 16(1), 117–125. https://doi.org/10.1038/s41592-018-0234-5

Pfau, T., Fiske-Jackson, A., Rhodin, M. (2016). Quantitative assessment of gait parameters in horses: Useful for aiding clinical decision making? Equine Veterinary Education, 28(4), 209–215. https://doi.org/10.1111/eve.12372

Ren, S., He, K., Girshick, R., Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In Advances in Neural Information Processing Systems (Vol.

2015-January). Retrieved from arXiv:1506.01497v3

Rooney, J. R., Thompson, K. N., Shapiro, R. (1991). A contribution to the study of velocity, stride length, and frequency in the horse. Journal of Equine Veterinary Science, 11(4), 208–209. https://doi.org/10.1016/S0737-0806(06)80978-0

Rose, N. S., Northrop, A. J., Brigden, C. v., Martin, J. H. (2009). Effects of a stretching regime on stride length and range of motion in equine trot. The Veterinary Journal, 181(1), 53–55. https://doi.org/10.1016/j.tvjl.2009.03.010

Serra Bragança, F. M., Broomé, S., Rhodin, M., Björnsdóttir, S., Gunnarsson, V., Voskamp, J. P., … Hernlund, E. (2020). Improving gait classification in horses by using inertial measurement unit (IMU) generated data and machine learning. Scientific Reports, 10(1). https://doi.org/10.1038/s41598-020-73215-9

Serra Bragança, F. M., Rhodin, M., van Weeren, P. R. (2018). On the brink of daily clinical application of objective gait analysis: What evidence do we have so far from studies using an induced lameness model? The Veterinary Journal, 234, 11–23. https://doi.org/10.1016/j.tvjl.2018.01.006

Vinika Gupta. (2021). Equine Gait Analysis, Body Part Tracking using DeepLabCut and Mask R-CNN and Biomechanical Parameter Extraction. [Master's Thesis, Auburn University] Auburn University, Auburn. https://etd.auburn.edu//handle/10415/7894

Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. (2019). Detectron2. https://github.com/facebookresearch/detectron2

Zimmermann, C., Schneider, A., Alyahyay, M., Brox, T., Diester, I. (2020). FreiPose: A Deep Learning Framework for Precise Animal Motion Capture in 3D Spaces. https://doi.org/10.1101/2020.02.27.967620

CHAPTER 3. PHENOTYPING OF ARCHITECTURE TRAITS OF LOBLOLLY PINE TREES USING STEREO MACHINE VISION AND DEEP LEARNING: STEM DIAMETER, BRANCH ANGLE, AND BRANCH DIAMETER

Highlights

- A novel investigation in pine tree architecture analysis by leveraging stereo machine vision and deep learning.
- An image processing pipeline to measure stem diameter, branch angle, and branch diameter under field conditions.
- A rapid, cost-effective, and automated tool for loblolly pine breeders to select superior genotypes in progeny tests.

3.1 ABSTRACT

Loblolly pine is one of the most planted forest tree species in the Southern United States for sawtimber production. The sawtimber yield potential of a pine tree is significantly impacted by its stem and branch architecture and are important traits in tree breeding programs. However, phenotyping these traits in the upper crown of pine trees is currently based on subjective visual assessments. This study investigated the feasibility of quantifying stem diameter, branch angle, and branch diameter of six-year-old loblolly pine trees in a progeny test using stereo 3D imaging, deep learning-based organ segmentation, and image and point cloud processing techniques. Instance segmentation of branches and stems was performed as well as principal component analysis (PCA) in 2D images, followed by 3D reconstruction of the segmented organs. The resulting 3D point clouds were further processed using random sample consensus (RANSAC) and statistical outlier removal to extract stem diameter, branch angle, and branch diameter. When compared to the manual measurements, the three system-derived parameters achieved RMSEs of 0.055 m, 5.0 deg, and 5.6 mm, respectively. In addition, Bland-Altman analyses showed that the stem diameter and branch angle estimations were found with limits of agreement of ±0.098 m and ±9.8 deg, respectively, with nonsignificant biases. On the other hand, branch diameter estimation showed -12.1 mm and 9.3 mm for lower and upper limits of agreement with a bias. The proposed system demonstrated promising potential as a high-throughput precision phenotyping tool for the characterization of loblolly pine tree architecture in field tests, facilitating the selection of superior genotypes with improved sawtimber properties.

Keywords. 3D Reconstruction, Branch Angle, Deep Learning, Loblolly Pine, Stem Diameter, Stereo Matching, Tree Architecture.

3.2 INTRODUCTION

Due to large-scale plantings and natural regeneration from 1930 through the turn of the century, loblolly pine (*Pinus taeda*) has become the most widely planted timber species in the Southern United States (Schultz, 1999). Tree improvement programs commenced at land grant universities and at the USDA Forest Service in the 1950s to provide improved genetics for the forest industry (Zobel and Talbert, 1984). Initially, superior-looking trees were selected from natural stands that in turn produced superior progeny. Repeated rounds of selection over successive breeding cycles led to the continued improvement of commercially important growth traits that include stem volume and stem form (McKeand et al.,2019). Decades of tree improvement, combined with increasingly sophisticated silvicultural practices, led to extensive plantations of improved loblolly pine driven by a desire to increase productivity, disease resistance, seedling survival, and shorten harvest rotations, leading the Southern U.S. to become the world's most productive timber region (Allen et al., 2005; Fox et al., 2007).

Sawtimber quality is largely determined by lumber volume and grade yield. Lumber volume is positively influenced by sawlog diameter (Steele et al., 1984). Lumber grade yield is elevated by several desirable log attributes that include a straighter stem with smaller and fewer branch knots present (SPIB, 1994). Thus, sawtimber quality is strongly characterized by stem form and by growth and development of the branches. For high quality timber production, a tall straight tree with little taper that has thin, flat-angled branches is considered a desirable tree with small knots (Ehrenberg et al., 1970). In progeny tests, the accurate assessment of sawtimber quality has proved challenging given the informative traits such as branch diameter and branch angle are typically measured using a subjective score (Cumbie et., al 2012). The direct measurement of these traits using a caliper and protractor is prohibitively expensive and working at heights within the tree crown would generate significant safety concerns.

For the objective assessment of tree architecture traits, there are two non-contact methods that have been described in the literature, namely range sensing- and image-based techniques. Range sensing-based methods such as LiDAR (light detection and ranging) and terrestrial laser scanning (TLS) are the most popular techniques for digitalizing individual tree architecture (Gorte et al., 2004; Kankare et al., 2014; Liang et al., 2016; Pfeifer et al., 2004; Srinivasan et al., 2014). Although using LiDAR can properly rebuild tree architecture (Tienaho et al., 2022), the collection of data with LiDAR is time-consuming, has low spatial resolution, and is expensive (Wilkes et al., 2017). Additionally, TLS produces large data that requires experts to process, which can be costly. Thus, with the existing tools, it remains a challenge to perform a high throughput pine tree architecture analysis in a progeny test environment due to their high planting density and occlusion by branches and needles.

As for image-based methods, despite the widespread use of 2D methods in image-based phenotyping techniques, they are not able to reflect three-dimensional (3D) quantities accurately (Gibbs et al., 2018; Kaminuma et al., 2004). There is evidence that analyzing plants in 3D space offers higher accuracy and robustness (Apelt et al., 2015). In order to perform phenotyping studies on plants, 3D imaging is essential to tracking the exact geometry of the plant (Ziamtsov and

Navlakha, 2020). 3D imaging not only provides quantitative descriptions, but also allows measurements of more traits than 2D images (Paulus et al., 2014). However, there is limited research that employs image-based approaches to reconstruct 3D models of pine trees under field conditions. On the other hand, for a variety of other plant species, recent advancements have been made in camera-based multi-view stereo vision (MVS). By combining more than two images, stereo machine vision is one of the passive imaging techniques used to reconstruct 3D canopy models. Using this method, a plant canopy's 3D reconstruction can be achieved with information such as height, width, volume, and leaf cover area (Nguyen et al., 2016). Reconstruction of 3D canopy models under field conditions has been successfully carried out using this method (Bao et al., 2019; Salas-Fernandez et al., 2017). Klodt and Cremers (2015) used a multiple-view imaging system to reconstruct 3D models of plants. Similarly, Li and Tang (2017) developed a low-cost 3D plant morphological trait analysis system for corn seedlings by combining two webcams and a time-of-flight (ToF) camera. This study shows that a high spatial resolution 3D reconstruction of the scene can be achieved using a low-cost setup. Inspired by the aforementioned studies in high-throughput plant phenotyping and with the recent developments in commercial RGB-Depth cameras, a 3D reconstruction of a scene can be achieved with a one-time factory calibrated commercial binocular camera (Paul et al., 2020).

It has been shown that deep learning-based object detection approaches can be used in plant and fruit segmenting and counting, including apple, orange, and other fruits (Chu et al., 2021; de Melo et al., 2020; Ganesh et al., 2019; Yikun Liu et al., 2021). Additionally, plant stems and leaves were detected and segmented using convolutional neural networks in some studies for sorghum and other canopies (Baweja et al., 2018; Sodhi et al., 2017; Xiang et al., 2021). A successful segmentation of forest tree stem and branches can potentially be achieved using deep learning models such as the work by Zhang et al. (2017) in detection of branches in apple trees using an R-CNN network.

This study aimed to assess the feasibility of using stereo machine vision, deep learning-based branch and stem segmentation, and image and point cloud processing techniques to quantify branch and stem characteristics of loblolly pine trees in breeding programs. The specific research objectives were to 1) develop an image processing pipeline to measure stem diameter, branch angle, and branch diameter, and 2) evaluate the performance of the pipeline in measuring the aforementioned traits at different tree height levels.

3.3 MATERIALS AND METHODS

3.3.1 *Acquisition of Ground-Truth Measurements*

Ground truth measurements and stereo image data were collected from 10 randomly selected 6-year-old loblolly pine trees growing in a North Carolina State University Tree Improvement Program progeny test located near Moultrie, Georgia. Flagging tape was used to demarcate 1 m intervals along the stem of each tree to a height of 6 m, as shown in Figure 3-1(a). Stem diameter was measured at each meter interval using a diameter tape. Three dominant branches were flagged within each of five intervals, with the first interval representing 1 - 2 m height above ground level, and the fifth interval representing 5 - 6 m above ground level. The diameter at the base of each flagged branch was measured using a digital caliper with a resolution of 0.1 mm. As shown in Figure 3-1(b) and Figure 3-1(c), branch angle relative to the vertical stem was measured using a digital angle finder with a resolution of 0.1 degrees. The flagging tape on each measured branch was color-coded so that each branch could be readily identified following the stereo image data collection.

(a) (b) (c)

Figure 3-1. (a) In-field acquisition of ground-truth data: (a) 1 m intervals up to a height of 5m were demarcated on the stem using flagging tape, and the stem diameter at each meter interval was measured using a diameter tape, (b) A digital angle finder was used to measure branch angle relative to the vertical stem, and (c) a digital caliper was used to determine branch diameter adjacent to the stem.

3.3.2 *Stereo Image Data Collection*

Stereo images of individual pine trees were collected using a ZED2 stereo camera (StereoLabs, France) mounted on a tripod (Fig. 3-2). Four stereo image pairs were collected every 15 deg, stating from horizontal and ending at a 45 deg from the horizontal line as shown in Figure 3-2(a). The resolution of the collected images was 2208×1242 pixels. The stereo camera had an angular field of view of 110° by 70°. The camera was initially pointed at the breast height level to capture the tree base, and then it was tilted upward to capture the entire tree. To train and test an instance segmentation model for branch and stem, 13 additional stereo image pairs were collected at different angles between the previous 4 angles randomly. These stereo pairs were collected from the 10 flagged trees. Four perspectives, 90 degrees apart (i.e., 45 deg, 135 deg, 225 deg, and 315 deg), around the tree were selected and the procedure of data collection was repeated for all perspectives as shown in Figure 3-2(b). In all cases, the stereo camera was kept 1.5 m from the tree, while the height of the camera was maintained at 1.30 m above ground level. The weather was sunny on the collection day. No artificial lighting was used during the collection of these data.

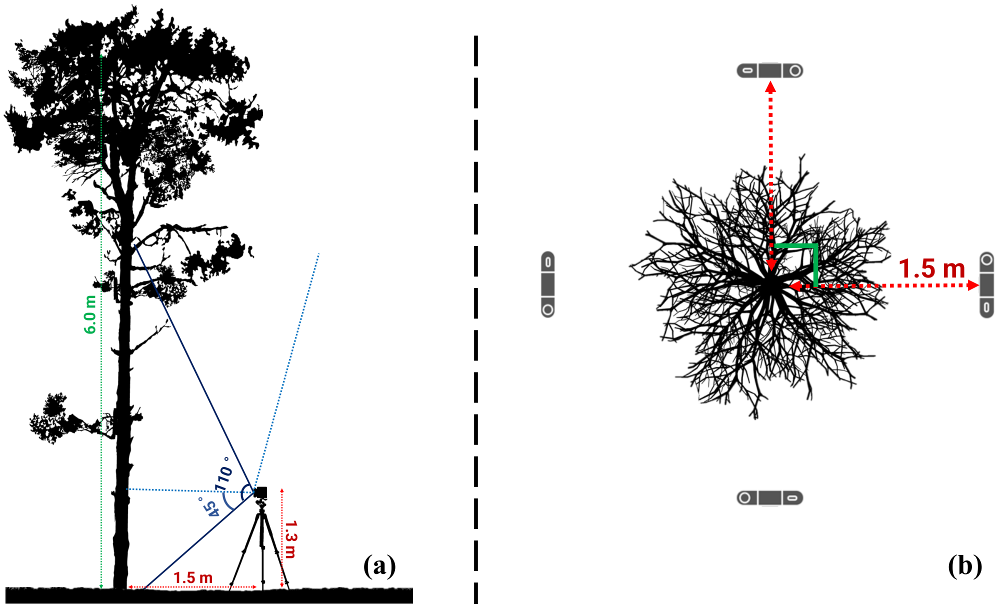


Figure 3-2. Stereo image acquisition setup. (a) A stereo camera has a 110-deg vertical angular field of view. To capture the entire tree architecture in multiple images, the camera was tilted on a tripod from 0 to 45 deg with respect to the horizontal plane, shown from side view. (b) The stereo images were also collected from 4 different perspectives that were 90 deg apart, shown from top view.

3.3.3 *Data Processing Pipeline*

A data processing pipeline was developed to detect the stem and branches of a tree in the center of stereo images, and further measure stem diameter, branch diameter, and branch angle (Fig. 3-3). Among its components are an instance segmentation model that detects and isolates branches and stems in images, and a stereo matching algorithm for generating 3D point clouds of the segmented organs. Additionally, a series of point cloud processing algorithms were utilized to compute the branch diameters and angles. The details are discussed in the following sections.

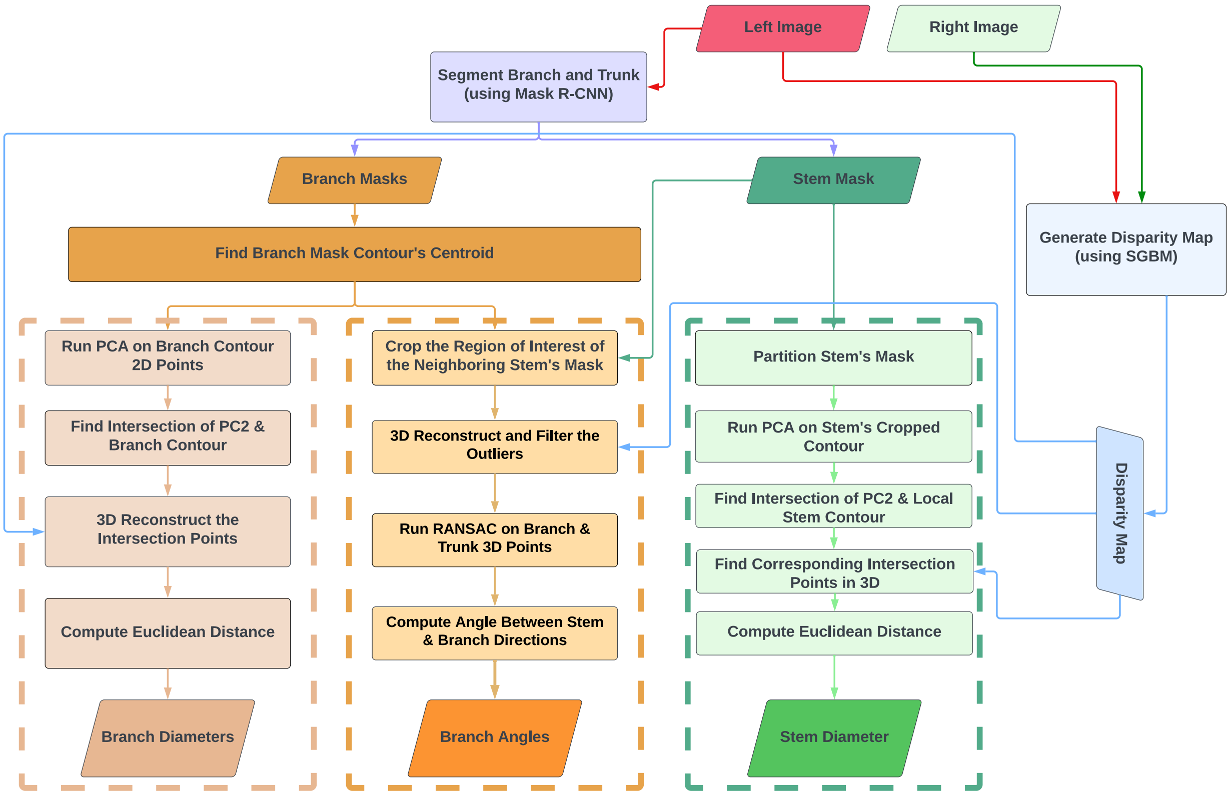


Figure 3-3. Data processing pipeline consisting of stem diameter, branch angle, and branch diameter estimation algorithms.

3.3.4 *Stem and Branch Instance Segmentation*

A polygon was drawn around the objects of interest, i.e., branches and stems using COCO Annotator (V0.11.1) (Brooks, 2019), as shown in Figure 3-4. The dataset included two classes: branch and stem. The 510 frames were randomly selected from the collected dataset. Since they were used for model training and testing of the segmentation model, these images were not part of the dataset that was collected for evaluation of the estimation of tree traits. There were 510 annotated stems and 9755 annotated branches in this dataset. The tree in the center of the image was annotated. Annotations were made for each branch based on its nearby stem diameter. Specifically, a branch was annotated such that its length was approximately equal to the stem diameter measured at the base of the branch. Stems on the other hand were thoroughly annotated.

This step was done because the branches were not straight at all times and if their curves were detected by the detection algorithm, the angle estimation algorithm may possibly produce subpar results. Lastly, the annotations were saved to a JSON file in COCO format.

Figure 3-4. Stem and branch annotations, stem indicated with a blue mask and the branches with orange ones.

A Mask R-CNN model that was pre-trained on the COCO dataset was fine-tuned from Detetctron2 (Wu et al., 2019) using the created dataset. This deep learning model generates high-quality segmentation masks for each instance in an image while efficiently detecting objects. A Resnet-101 C4 3X backbone was used in this model (Lin et al., 2014). 70% of the annotated images were randomly selected for model training and the remaining 30% were used for testing. A comprehensive specification of the dataset is shown in Table 3-1. Training parameters included a learning rate of 0.001 with a decay rate of 0.05, 200 warm-up iterations, and 500 maximum iterations. The batch size was 2, and regions of interest (ROI) in images were 128.

Table 3-1. Summary of the datasets and model for instance segmentation of stem and branch.

Model	Backbone	Pre-trained	Categories	Total	Tool	Training	Testing
Mask R-CNN	ResNet -101 C4 3X	COCO Dataset	Stem	510	COCO Annotator	70% of Total	30% of Total
			Branch	9755			

3.3.5 *Stereo 3D Reconstruction of Stems and Branches*

ZED software development kit (SDK) was used to acquire the stereo camera parameters. The stereo images were rectified during collection procedure by the camera. Semi-global stereo block matching (SGBM) in the OpenCV library was then used to generate the disparity map (Bradski, 2000). SGBM has demonstrated the capacity to manage textureless areas with a smoothness constraint in its objective function (Hirschmuller, 2008). The matching block size and the number of disparities were chosen to be 19 pixels and 112 pixels, respectively, considering the image resolution, camera-to-tree distance, and the 0.12 m baseline of the stereo camera. Finally, the stereo camera parameters and disparity map were used to project the 2D pixels back into 3D space using Equation 3-1:

$$\begin{cases} x = \dfrac{u - c_x}{W} \\ y = \dfrac{v - c_y}{W} \\ z = \dfrac{f}{W} \\ W = \dfrac{d}{T_x} \end{cases} \tag{3-1}$$

where x, y, and z are the coordinates of a point in 3D space, u and v are the 2D image coordinates of a pixel in the left image, W is a dummy variable, d is the disparity value of the pixel, as well as the intrinsic parameters of the ZED2 stereo camera including the coordinates of the principal point of the left camera of the stereo camera in the image coordinate system (C_x = 1136.7 pixels and C_y

= 659.75 pixels), focal length (f = 1058.75 pixels), and baseline between the two cameras (T_x = 0.12003 m).

3.3.6 *Stem Diameter Estimation*

The following method was developed to estimate stem diameters. For each of the left-view images, the stem mask detected by the Mask-R-CNN model was used to segment stem pixels and generate a bounding box. Subsequently, a sliding window of 200 pixels tall inside the bounding box was examined every 75 pixels along the vertical axis to crop out sections of the stem mask as shown in Figure 3-5(a). The window height was chosen as approximately twice the average stem diameter.

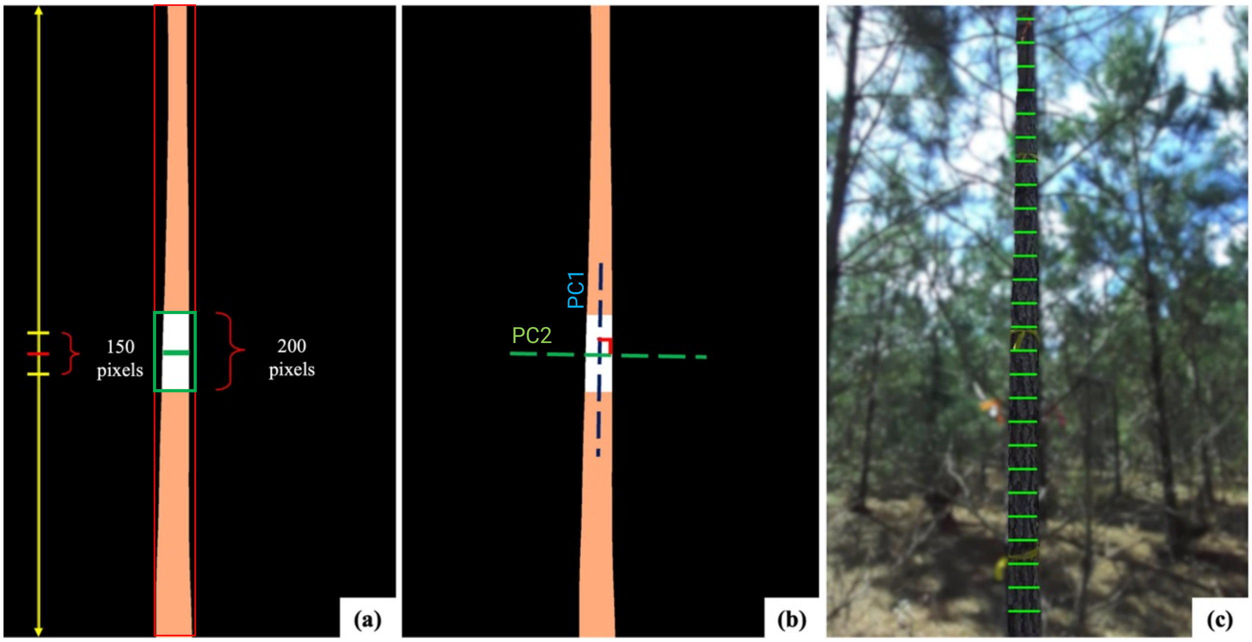


Figure 3-5. A procedure for estimating stem diameter involves: (a) determining the precise location where the stem's diameter should be computed and their corresponding windows to crop the stem. (b) As a second step, the PCA is run on the cropped part of the stem (shown in white) to determine the first and second eigenvectors of the points within the section of interest, (c) Identify all the points resulting from the intersection of the second eigenvectors with the stem's contour along the tree, which are 75 pixels apart.

The *PCACompute2* function from the OpenCV library (Version: 3.14.16) (Bradski, 2000) was then used to conduct a principal component analysis (PCA) on each cropped stem section and find the local stem orientation as the first eigen vector in 2D image space. Next, the centroid of the cropped stem section and the second eigenvector (PC2) were used to create a line perpendicular to the local stem orientation. The two points at which this line intersected with the contour of the stem mask were considered as the references for calculating stem diameter along the two edges of the stem in 3D space, as shown in Figure 3-5(b). Then, using the *remove_statistical_outlier* function from the Open3D library (Version: 0.15.0), with a manually defined number of neighbors (*nb_neighbors* = 20) and a standard ratio (*std_ratio* = 2), the outliers of the 3D point cloud were filtered out. The first parameter specifies how many neighbors should be considered when calculating a point's average distance, and the second determines the threshold level according to the standard deviation of the average distance between points in the point cloud. Filtering is more aggressive with a lower number of standard ratios. Subsequently, three pixels up and three pixels down in the direction of the found PC1 were located around each intersection point. Then, the corresponding 3D coordinates of the 14 points were retrieved from the 3D point cloud. The centroid of the seven points on each side of the stem was calculated. Lastly, the diameter of that section of the stem was then computed using the Euclidean distance between the two found centroids in 3D space.

The algorithm was repeated for the entire stem, from the bottom to the top, as shown in Figure 3-5(c). As a result, a stem diameter value was associated with every picked height. to find the stem diameter at the tags, the 2D locations of the right side of the tags were annotated in a CSV file for all the left images in the dataset. The diameter was determined for each tag's coordinate by interpolating the two nearby values.

3.3.7 *Branch Angle Estimation*

The trained Mask R-CNN model was used to extract branch and truck masks, which were then exploited further to estimate branch angle. For each branch mask, a local region of the stem was cropped by placing a window as described in the previous section at the same height of the branch mask centroid (Fig. 3-6). A line was fitted to the 3D reconstructed points of each branch mask using random sample consensus (RANSAC) (Fischler and Bolles, 1981). Furthermore, as shown in Figure 3-6, RANSAC was used to fit a line to the adjacent stem points in 3D space. The *remove_statistical_outlier* function in Open3D library was applied to the point clouds of the stem and branches before the lines were fitted to them. The results of the preceding procedures were the directions of the two extracted lines. It was then necessary to check the direction of the fitted lines. If the stem or branch lines were pointing downwards, they were negated. This was done considering the fact that the branches in the dataset had an angle between 20 and 80 degrees. The branch angle was determined using Equation 3-2.

$$\theta = \cos^{-1}\frac{(\vec{\mathrm{U}} \, . \, \vec{\mathrm{V}})}{|\vec{\mathrm{U}}||\vec{\mathrm{V}}|} \tag{3-2}$$

where θ is the angle between the two vectors of $\vec{U}$ and $\vec{V}$, which are the direction vectors of the lines fitted to the branch and stem points, respectively.

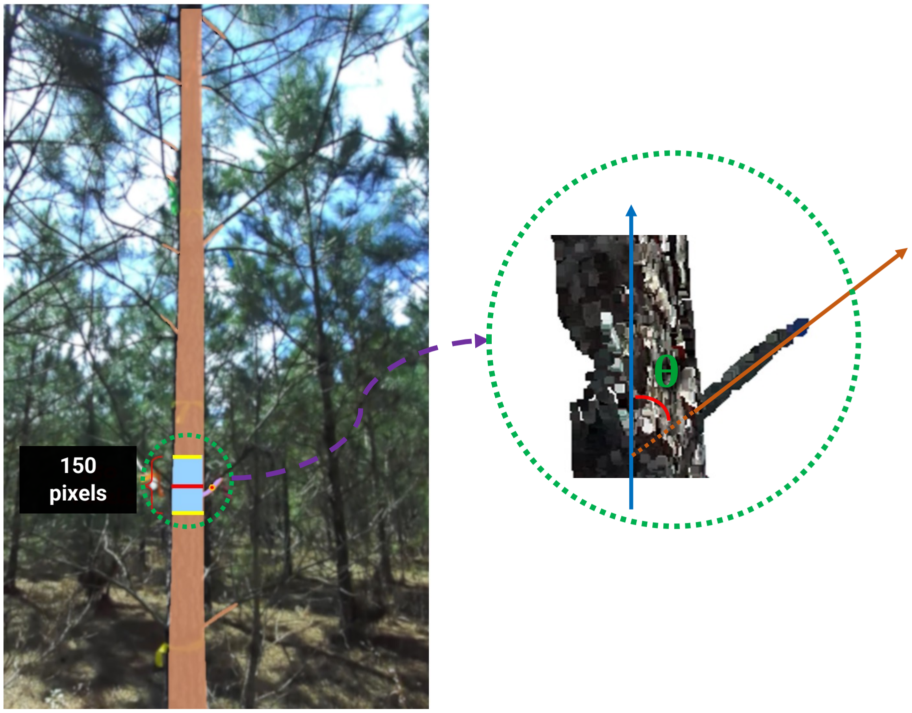


Figure 3-6. Branch angle estimation procedure. The branch angles were calculated using RANSAC by fitting a line to the filtered 3D points of the branch mask and another line to its local stem. The directions of the fitted lines were assessed and corrected to point to the sky.

3.3.8 *Branch Diameter Estimation*

Using *findContours* from the OpenCV library, a contour was created for each branch mask. Next, a PCA analysis was performed on the contour of each branch mask using the *PCACompute2* function in OpenCV. The second principal component, in this case, reflects the direction perpendicular to the branch. The intersection points of the derived PC2 direction and the contour of the branch were found. A line from the centroid of the derived branch mask and in the direction of the second principal component (PC2) was drawn. Backprojecting 2D points into 3D space was achieved by using a disparity at the centroid of the branch contour. Due to the fact that branches may not always be frontoparallel to the camera, the base of a branch that appears in an image might not always be the actual base of the branch. It is possible that this phenomenon can introduce discontinuities into the extracted disparity map and increase the error in branch diameter estimation. Therefore, the contour's centroid was picked instead of the base in the images. In the images, it can be assumed that the disparity was the same along the line perpendicular to the branches because the branches were thin and there were a few pixels per branch. The found intersection points were in the neighborhood of 2 to 5 pixels of the branch centroid and because of the previously mentioned reason, the disparity were considered to be the same in these points. This property allowed projecting the two 2D points into 3D space using the disparity at the branch centroid. As shown in Figure 3-7, the Euclidean distance between two points in 3D space was used to determine branch diameter. Furthermore, if some branches were found multiple times by the trained Mask R-CNN model, the system could detect them by examining their distance from each other. If the centroids of two branch masks were closer than 10 pixels, they were considered the same. And the values for branch angle and diameter of those close branches were averaged.

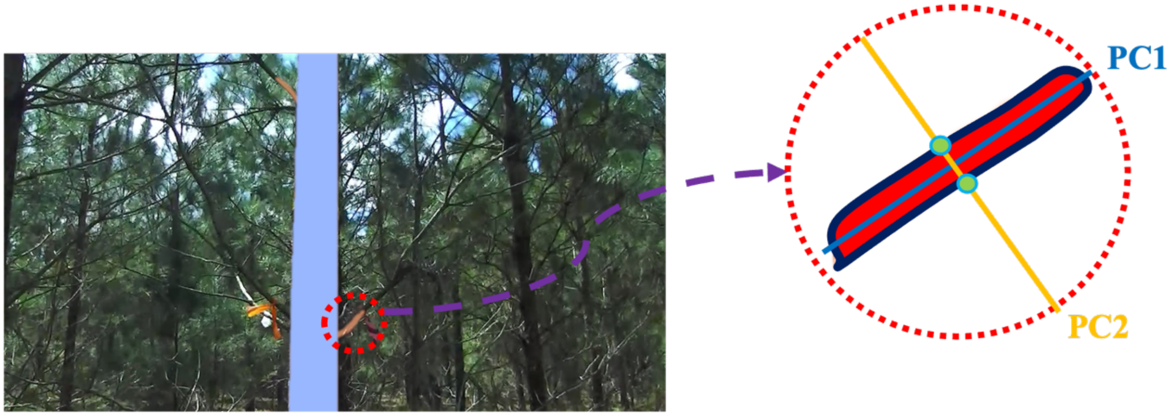


Figure 3-7. The procedure for estimating the diameter of a branch is carried out by finding the intersection points between a line drawn from the centroid of the branch mask using second principal component direction (PC2), and the contour itself, back projecting these two points into 3D space, and then computing the Euclidean distance between the two points.

3.4 EVALUATION METHODS

3.4.1 *Instance Segmentation Evaluation*

Average precision (AP) was used to evaluate the accuracy of the Mask R-CNN instance segmentation model for stem and branch. AP was calculated by integrating the area under the precision-recall curve, with higher values meaning higher accuracy. Two AP values were computed for "Stem" and "Branch", respectively. The performance of Mask R-CNN in the instance segmentation task was assessed in test datasets.

3.4.2 *Evaluation for Architecture Traits*

Since four sides of each tree were imaged with four imaging angles per side, the same instance of stem or branch could be detected from multiple images. For the same stem or branch, a trait of interest was estimated for each detected instance and all estimations were averaged to compare with the infield measurement. Root mean squared error (RMSE) and mean absolute error (MAE) were used to assess the accuracy of stem diameter, branch angle, and branch diameter estimations. RMSE was calculated using Equation 3-3.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{\left(\hat{Y}_i - Y_i\right)^2}{n}} \tag{3-3}$$

where $\hat{Y}$ is the computed value for either stem diameter, branch angle, or branch diameter, Y is the in-field measurement for the aforementioned values, n is the total number of the measurements, and finally, i is the present measurement. Similarly, to calculate RMSE, Equation 3-4 was utilized. The parameters are the same as Equation 3-3.

$$MAE = \frac{\left|\sum_{i=1}^{n}\left(\hat{Y}_i - Y_i\right)\right|}{n} \tag{3-4}$$

Bland-Altman (B-A) analysis (Bland and Altman, 1999) was employed to evaluate bias and limits of agreement (LoAs) between in-field manual measurement (Y) and system-derived measurement ($\hat{Y}$). The B-A analysis can be visualized in a scatter plot (i.e., B-A plot), where x-axis is the mean of the two measurements and y-axis is the difference of the two measurements (Eq. 5). The bias is quantified by the mean of the differences between the two measurements (Eq. 6), and the standard deviation (SD) measures the dispersion of the differences (Eq.7). Then, the upper and lower LoAs are quantified as bias plus and minus 1.96 SD, respectively (Eq. 8 and 9) at a 95% confidence interval. In comparison with the reference method, the closer the bias is to zero, the more accurate the proposed method is. If the LoAs are close to the bias line and the spread of differences is very small, the standard deviation is small. A linear regression was also performed in each B-A plot and the significance of the slope indicated whether the bias was constant across measurements (Bosch et al., 2018; Hatrisse et al., 2022).

$$Difference = Reference\ Method - \ Proposed\ Method \tag{3-5}$$

$$Bias = Mean\ (Difference) \tag{3-6}$$

$$SD = \ Standard\ Deviation(Difference) \tag{3-7}$$

$$LimitAgreementHigh\ (Upper\ LOA) = Bias + 1.96\ \times\ SD \tag{3-8}$$

$$LimitAgreementLow\ (Lower\ LOA) = Bias - \ 1.96\ \times\ SD \tag{3-9}$$

3.4.3 *Experiment Environment*

A Linux Ubuntu 20.04.4 LTS workstation was used to train and fine-tune the Mask R-CNN model for branch and stem segmentation, as well as to analyze data. Furthermore, the system was powered by a Ryzen Threadripper 2970WX 24-core AMD processor, 64 GB of RAM, and two NVIDIA Titan RTX GPUs.

3.5 RESULTS

3.5.1 *Stem and Branch Instance Segmentation*

Figure 3-8 shows losses for object classification, bounding box regression, and mask segmentation tasks during training of the Mask R-CNN model. The classification accuracy started from ~0.25 and approached ~0.9. For the test dataset, the Mask R-CNN model yielded AP values of 28.25 and 39.54 for the detected branch and stem instances, respectively. The mask segmentation loss decreased from 2.0 to approximately 0.25 at the 250th iterations and became relatively stable. As expected, the same is true for classification loss. The classification accuracy on the contrary had an increasing trend, approaching 0.9.

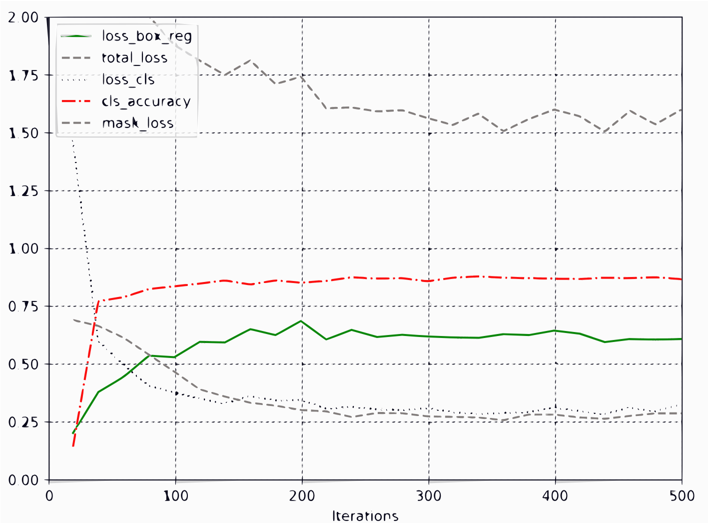


Figure 3-8. Training losses and classification accuracy for training of Mask R-CNN model for stem and branch instance segmentation. Regression loss of bounding box is shown by loss_box_reg, total loss by total_loss, classification loss by loss_cls, accuracy of classification by cls_accuracy, and loss of mask detections by mask_loss.

Qualitative examples of stem and branch instance segmentation from the test dataset are shown in Figure 3-9. In this figure, the trained model is illustrated with common success and failure cases. Each image contained a fully detected stem. And only trees in the front of the scene and in the center of the image were detected by the model. The model failed to detect some branches in some cases, while it detected others multiple times.

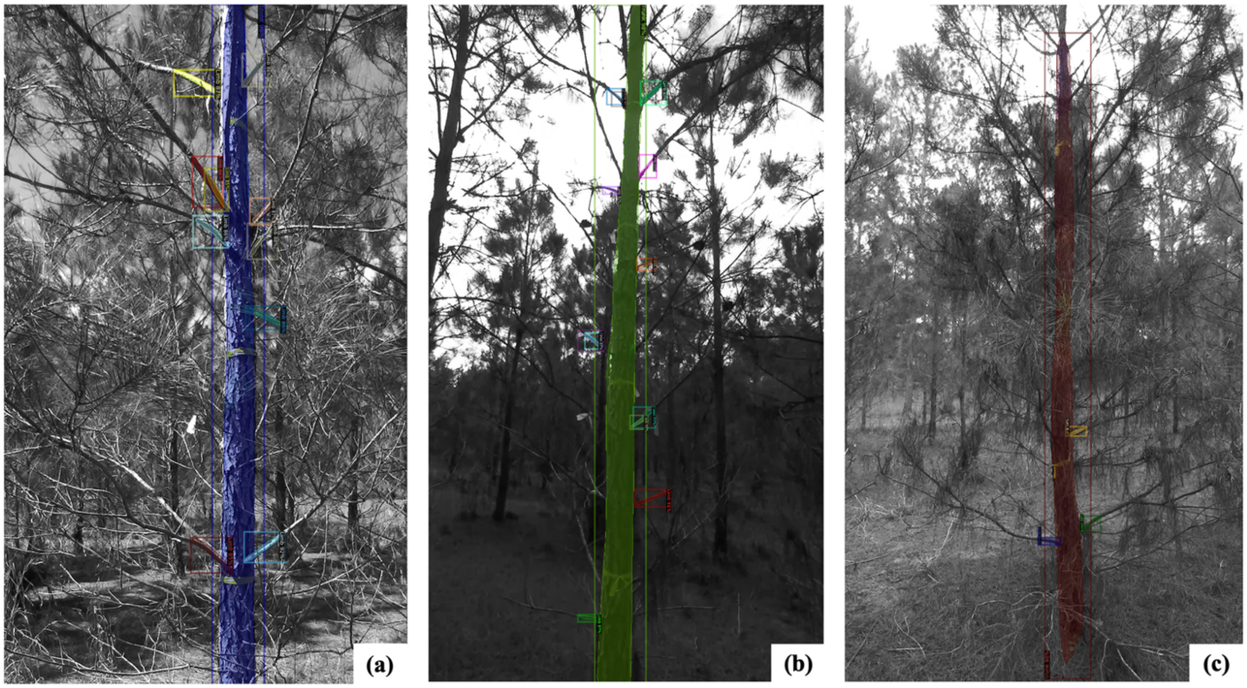
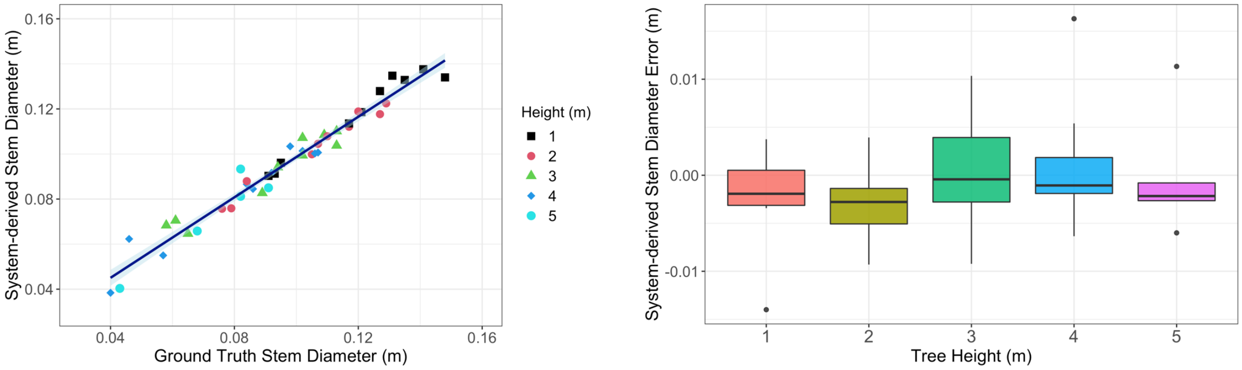


Figure 3-9. Masks and bounding boxes of detected stem and branch instances by a Mask R-CNN model. (a) Almost all the branches and the whole stem are detected and segmented. (b) Some branches are detected more than once. (c) Some are not detected due to being occluded by dense needles. No background trees were detected. The background was separated from the detections for better illustration.

3.5.2 *Stem Diameter Estimation*

Average RMSE and MAE for the stem diameter calculations were 0.055 m and 0.004 m, respectively, with 45 successful computations out of 50 measurements in the field. Figure 3-10(a) shows the linear regression model between ground truth measurements and system-derived stem diameters, with a slope of 0.89, an intercept of 0.009 m, and an R^2 of 0.96. An error box plot in Figure 3-10(b) reveals that the first section of the tree had the smallest range of error, while the third sections had the largest range of error. The range of error in the fifth section (5 to 6 m) appears to be smaller than that of the first section but this is due to the lower number of samples.



(a) (b)

Figure 3-10. Stem diameter estimation results. (a) System-derived stem diameter estimation against in-field measurement at different heights up the stem with a linear regression model fitted. (b) Boxplots for stem diameter estimation errors at different height levels. The lower and upper sides of a box show the 25th and 75th percentiles, respectively. The horizontal line inside a box represents the median. The two whiskers on either side of the middle 50% line representing scores outside the middle, including the min and max and excluding the outliers.

According to Table 3-2, the standard deviation of error increased from 0.004 m in the lowest height to 0.006 m in the highest. At different height levels, the stem diameter estimation algorithm had an RMSE of 0.004 m and a MAE of 0.003 m for height level of 1 m and an RMSE of 0.006 m and a MAE of 0.004 m for height level of 4 m as minimum and maximum errors. In all ten trees, successful system-derived measurements were consistently 100% except for the highest section, where 5 out of 10 cases could not be calculated because the pipeline, specifically the stereo matching algorithm, would fail generating results for this height.

Table 3-2. RMSE, MAE, the ratio of successful stem diameter estimations, and error standard deviation at different stem heights in the dataset.

Height Range (m)	**The Ratio of Successful Measurements**	**RMSE (m)**	**MAE (m)**	**Error Standard Deviation (m)**
1 – 2	10 / 10 (100%)	0.004	0.003	0.004
2 – 3	10 / 10 (100%)	0.004	0.003	0.003
3 – 4	10 / 10 (100%)	0.006	0.004	0.006
4 – 5	10 / 10 (100%)	0.006	0.004	0.006
5 – 6	5 / 10 (50%)	0.005	0.004	0.006
Total	45 / 50 (90%)	0.005	0.004	0.005

In Figure 3-11, the Bland-Altman analysis of stem diameter estimation shows a lower LoA of -0.009m and an upper LoA of 0.011m. Although the bias was 0.001 m, a t-test was performed to test whether the mean difference between the two measurements was different from zero and the resulting p-value of 0.2442 suggested that there was no statistically significant bias. According to Figure 3-11, the slope of the fitted regression line was 0.007 and statistically significant with a p-value of 0.032 for stem diameter. According to the nonzero slope, the greater the average stem diameter was, the greater the difference between the two measurements was. Also, from the positive slope it can be understood that the system-derived measurements went from overestimation in stem diameter computation to underestimation as the diameter increased. Furthermore, the positive slope can be explained by the fact that stem diameter computations were overestimated for below 90 mm average diameter and above that, the stem diameter was underestimated.

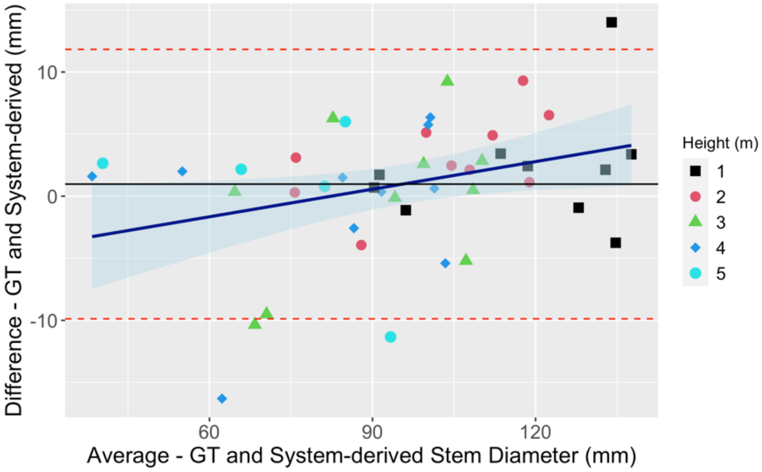
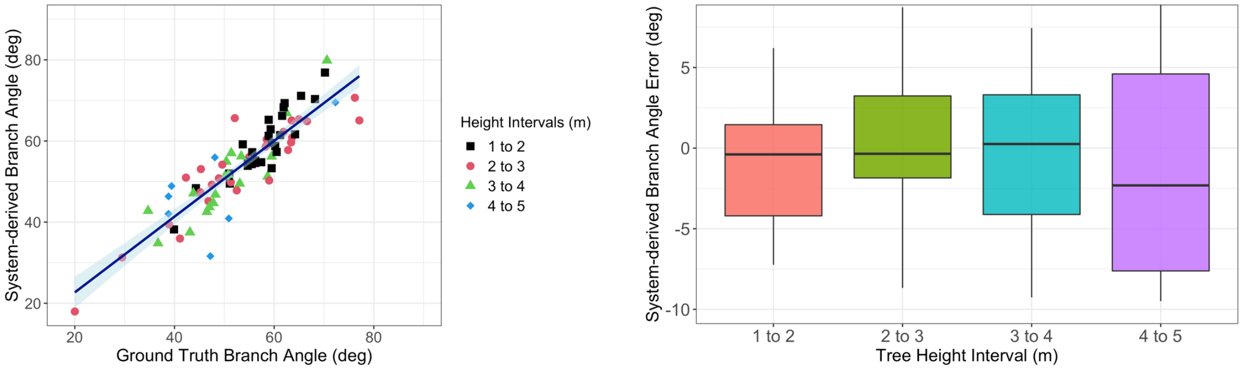


Figure 3-11. Bland-Altman analysis on system-derived stem diameter against ground truth (GT) measurements (i.e., in-field manual measurements). Upper and lower LoAs are shown as the dotted red lines, and bias shown as black line. A regression line was fit to the difference between the measurements and their average.

3.5.3 *Branch Angle Estimation*

The system-derived branch angles were evaluated against in-field measurements, and a linear regression model was fitted to the scatter plot shown in Figure 3-12(a), with a slope of 0.93, an intercept of 4.0 deg, and an R^2 of 0.78. The 4 - 5 m section of the trees had the largest error range, as illustrated in Figure 3-12(b), with only 8 out of 30 branch angle measurements being successful, as opposed to the 1 - 2 m section with 27 successful branch angle measurements. The results of the last segment were removed since the Mask R-CNN model was unable to detect the branches in the tallest section (i.e., 5 - 6 m) in any of the ten trees.



(a) **(b)**

Figure 3-12. Branch angle estimation results. (a) System-derived branch angle against in-field measurements at multiple height levels with a linear regression model fitted. (b) Boxplots for branch angle estimation errors at different height levels. The lower and upper sides of a box show the 25th and 75th percentiles, respectively. The horizontal line inside a box represents the median. The two whiskers on either side of the middle 50% line representing scores outside the middle, including the min and max and excluding the outliers.

The ratio of successful detection decreased when the branch instances were intended to be retrieved from a higher section on the trees, from 90% in the first section to 26.66% in the fourth. This fashion is congruent with the trends of RMSE and MAE for each section. The 4 - 5 m section had an RMSE of 7.9 deg and a MAE of 7.2 deg, and the first section had an RMSE of 3.5 deg and a MAE of 2.8 deg as described in Table 3-3. Total RMSE and MAE for branch angle estimation were 5.0 deg and 3.9 deg, respectively. Also, the standard deviation of error was 3.4 deg in the range height of 1 - 2 m and it increased to 9.0 deg at the highest level. The mean of standard deviation of branch angle estimation error was 5.0 deg.

Table 3-3. RMSE, MAE, the ratio of successful branch angle estimations, and error standard deviation at different stem heights in the dataset.

Height Range (m)	**The Ratio of Successful Measurements**	**RMSE (deg)**	**MAE (deg)**	**Error Standard Deviation (deg)**
1 – 2	27 / 30 (90%)	3.5	2.8	3.4
2 – 3	26 / 30 (86.66%)	5.1	3.9	5.3
3 – 4	18 / 30 (60%)	4.5	4.0	4.8
4 – 5	8 / 30 (26.66%)	7.9	7.2	9.0
Total	79 / 120 (65.83%)	5.0	3.9	5.0

The Bland-Altman analysis of the in-field and system-derived branch angle measurements shows a lower LoA of -10.3 deg, a bias of -0.4 deg, and an upper LoA of 9.4 deg, as shown in Figure 13. The p-value of the slope in this scatter plot (between difference and average of measurements) was 0.326, which indicates that the slope is zero. Furthermore, the mean of the difference between the measurements was tested with t-test and resulted in a p-value of 0.437, suggesting that bias is statistically 0.

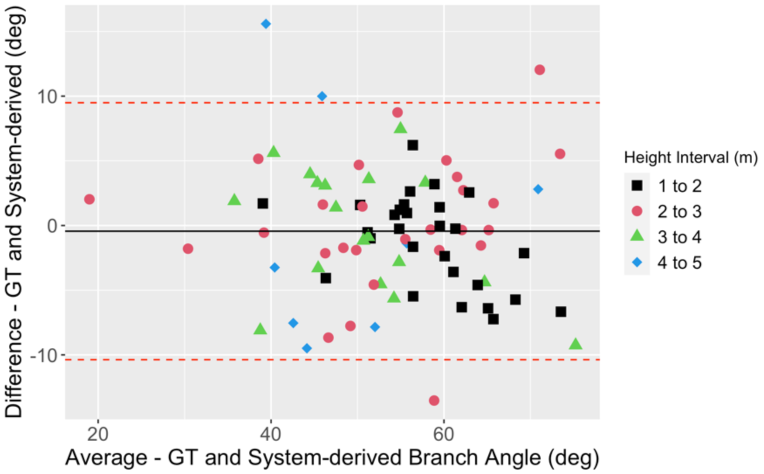
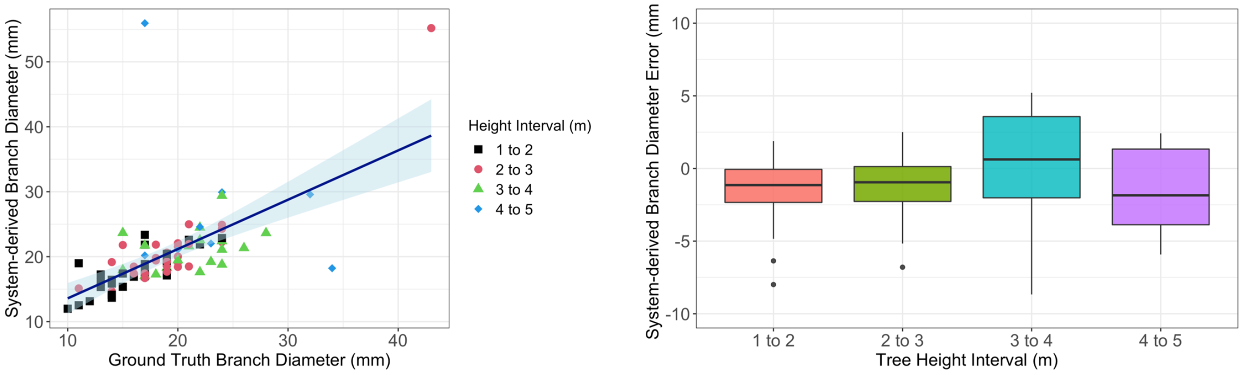


Figure 3-13. Bland-Altman analysis on system-derived branch angle against ground truth (GT) measurements (i.e., in-field manual measurements). Upper and lower LoAs are shown as the dotted red lines, and bias shown as black line.

3.5.4 *Branch Diameter Estimation*

Figure 3-14(a) displays the results of computing the branch diameters at each segment of the loblolly pine trees against manual measurements taken in the field. A linear regression model was fitted between the in-field measurements and the algorithm-derived estimations. The line's slope and intercept were 0.7 mm and 5.9 mm, respectively. As shown in Figure 3-14(b), the range of error increases with height, with smallest at the lowest section and the widest at the 4 m to 5 m section. Also, for the first two sections the system-derived values showed an overestimation of the measurements as well as the final section. On the contrary, the for the third section the system had an underestimation of values.



(a) (b)

Figure 3-14. Branch diameter estimation results. (a) show the system-derived branch angle versus in-field measurements for different tree sections with a linear regression model fitted, and (b) illustrates a series of boxplots for branch diameter estimation errors at different height levels. The lower and upper sides of a box show the 25th and 75th percentiles, respectively. The horizontal line inside a box represents the median. The two whiskers on either side of the middle 50% line representing scores outside the middle, including the min and max and excluding the outliers.

According to Table 3-4, the first section from the bottom of the tree had an RMSE of 2.6 mm and a MAE of 1.8 mm calculated with a 96.6% success rate, whereas the highest section had an RMSE and a MAE of 14.2 mm and 8.8 mm, respectively, with a success rate of 26.66%. In total, 82 out of 120 branches were detected to compute diameter. The RMSE and MAE of the total measurements were 5.6 mm and 2.9 mm, respectively. In line with previous results, the error standard deviation for the estimation of branch diameter increased from 2.2 mm at the lowest section to 15.5 mm at the highest section. The mean error standard deviation was 5.4 mm for all branch diameters.

Table 3-4. RMSE, MAE, the ratio of successful branch diameter estimations, and error standard deviation at different stem heights in the dataset.

Height Range (m)	**The Ratio of Successful Measurements**	**RMSE (mm)**	**MAE (mm)**	**Error Standard Deviation (mm)**
1 – 2	29 / 30 (96.66%)	2.6	1.8	2.2
2 – 3	27 / 30 (90%)	3.3	2.2	3.0
3 – 4	18 / 30 (60%)	3.6	3.0	3.8
4 – 5	8 / 30 (26.66%)	14.2	8.8	15.5
Total	82 / 120 (68.33%)	5.6	2.9	5.4

Figure 3-15 depicts a Bland-Altman analysis on the branch diameter measurements, showing lower and upper LoAs of -12.1 mm and 9.3 mm, respectively. The system-derived measurements showed a bias of -1.4 mm compared to manual measurements. The fitted regression line has statistically significant non-zero slope (slope = -0.302 with p-value = 0.007). The bias of the measurements was further analyzed by a t-test and the result with the p-value being 0.0224 suggested that there was a non-zero bias in the system-derived measurement when compared to the ground truth. In Figure 3-15, it can be seen that the measurements in the last section were highly deviated by the fact that the stereo matching algorithm failed to generate the correct disparity since the objects in the imagery were too small and far away.

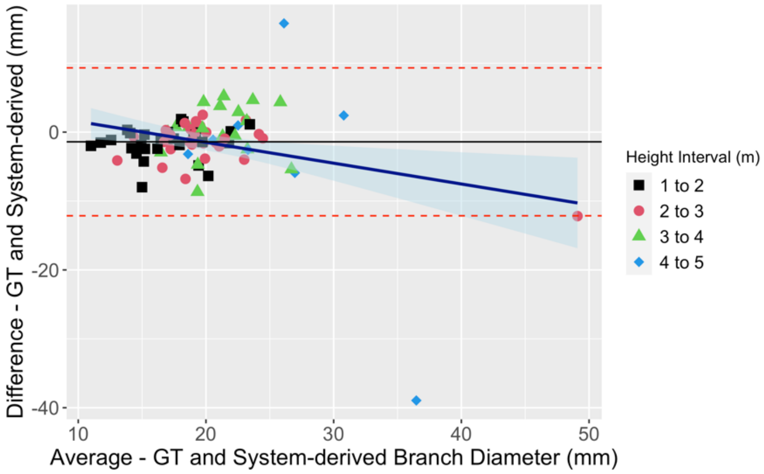


Figure 3-15. Bland-Altman analysis on system-derived branch diameter against ground truth (GT) measurements (i.e., in-field manual measurements). Upper and lower LoAs are shown as the dotted red lines, and bias shown as black line. A regression line was fit to the difference between the measurements and their average.

3.6 DISCUSSION

Overall, Mask R-CNN performed reasonably well for instance segmentation of stem and branch given that the images were acquired under challenging infield lighting conditions (i.e., low light and backlight). Between the two tree organs, stem instance segmentation achieved a far higher AP value than branch counterpart. This is mostly likely due to the large difference in size between stem and branch. It is well known that detection of small objects is challenging due to the lack of sufficient features for CNNs to extract (Nguyen et al., 2020). According to the "small object dataset" developed by Chen et al. (2017), an object is considered small when it has a mean relative overlap between the bounding box area and the image area between 0.08% and 0.58%. Branch masks in the dataset occupied between 0.1% and 0.3% of the whole image area. In addition, since each branch was annotated from the base to a somewhat arbitrary length, the Mask R-CNN model might detect the location of the branch accurately but predicted a mask of a different length. This could reduce the IoU value for each branch and thus overall AP. In contrast, each stem contour was completely annotated in each image. Stem instance segmentation typically did not fail.

However, branch instance segmentation could fail when occlusions by other branches and needles occurred or when branch in higher section of the stem became smaller in the image.

To the best of our knowledge, for the first time, a stereo machine vision system was used to measure a pine tree's stem diameter in an outdoor setting. Our method produced comparable results to existing remote sensing-based methods that have been used to measure stem diameter. Sun et al. (2022) combined a Kinect depth camera with deep learning and achieved a mean squared error of 0.0038 m for stem diameter estimation of apple trees (stem diameters ranging between 30 mm and 70 mm) at a certain height from an average distance of 2 m. Our proposed method for estimating the stem diameter has similar RMSE to the previously mentioned method and can be used to get measurements rapidly and accurately from different heights. Another study utilized a LiDAR-based method for stem segmentation and stem diameter estimation of boreal forests of Scot pines with an RMSE of 0.0039 cm for DBH (diameter range: 10 cm to 60 cm) (Kukkonen et al., 2022). Additionally, in the aforementioned work measurements were taken at a specific distance, angle, and location. Furthermore, LiDAR is widely considered one of the most expensive components in remote sensing. As opposed to 64 or 128 sparse rotating laser beams that LiDAR signals are inherently limited to, stereo cameras provide dense depth maps at much lower costs (Wang et al., 2019). Stereo matching and image processing algorithms allow the camera to function at a variety of angles and distances. As a result, users are no longer limited to modeling the morphological values of trees based on specific tree parameters such as height. Instead, they can obtain in-field measurements of different tree architecture traits. Prior studies only evaluated stem diameter at breast height, and extensive mathematical procedures were used to perform stem diameter estimation. The main issue in those studies is that they do not have diameter measurements along the tree stem. This could lead to an inaccurate estimation of stem volume because stem diameters of different genotypes could taper differently as a function of height (Gorte et al., 2004).

Traditionally, branch characteristics are visually assigned ordinal scores based on their appearance, which is subjective and inaccurate (Schomaker, 2007). It is not practical to measure the branch and stem traits at multiple heights using a protractor and caliper as there are typically hundreds of trees to be measured in a progeny test. Additionally, the scorer would need to be elevated to reach the branches thereby raising safety concerns. Using the method presented in this study, multiple branches are characterized in a high-throughput and safet manner by capturing one stereo image pair. This system mitigates safety risks and avoids cumbersome data collection procedures. The rapid stereo machine vision-based system is more accurate than subjective visual grading of branches by offering objective, quantitative results. Despite the branch angle estimation algorithm having lower and upper levels of agreement of -10.3 deg and 9.4 deg with branch angles ranging between 20 deg and 80 deg, the algorithm can be effectively used to categorize branch angles into two or three levels that are expected to have good correlation with knot size that impacts wood quality (Trincado and Burkhart, 2008). Outdoor phenotyping of individual tree branches has rarely been studied in the literature. Analyzing tree architecture and specifically branch morphological traits with 3D terrestrial laser scanning system in conjunction with model fitting methods was performed in some studies (Malhi et al., 2018; Raumonen et al., 2013). Although TLS provides a dense point cloud, in terms of technology, it is extremely expensive. As raw scanned models contain many small mistakes, they must be manually corrected to achieve high levels of detail (Radosevic and Rizvić, 2012). To the best of our knowledge, this study represents the first instance that stereo machine vision has been used for analyzing forest tree branch architecture. Using a turn table and a single camera a 3D point cloud was reconstructed by Lou et al. (2015) of various plant species (Arabidopsis, oat, wheat, maize, forage grass (lolium), clover, physalis, brassica, etc.) in an indoor setting. This approach had an RMSE of 1.88 deg for branch angle estimation. If the camera is moved around the trees in an outdoor setting, this method could potentially be used. However, it adds complexity to the process as multiple images need to be processed. As for our branch diameter estimation, the LoAs were large when compared to the range of measurements. The large errors were likely caused by the limitations of the stereo matching accuracy for small objects and thin structures (Ranftl et al., 2012). Additionally, because branches were not always parallel to the image sensor, the branch bases visible in the images may not be the actual branch bases. This creates a discontinuity at the branch bases and thus the branch diameter was measured distal from the branch base. This can potentially deteriorate the performance of the proposed pipeline in measuring branch diameter.

As for potential future direction of our developed system, quantification of stem morphology can potentially be made with the 3D point cloud of the stem. A conifer stem tilted from its normal vertical orientation inhibits the growth of the terminal shoot and induces compression wood to form on its lower side. Downgrading occurs due to compression wood, which is a severe sawtimber defect (Warensjö and Rune, 2004). Thus, stem straightness is of significant importance as a breeding objective. To increase the precision and the accuracy of the proposed pipeline, further improvements may be incorporated. A few modern stereo matching algorithms have recently been developed for thin objects. Several studies have attempted to more accurately handle stereo matching of thin objects (Dai et al., 2022; J. Li et al., 2022). Also, for improving binocular stereo vision system matching accuracy, some studies implemented stereo matching algorithms on images stored as RAW format (Liu et al., 2018). Furthermore, Chen and Jung (2018) used convolutional neural networks (CNNs) for improved stereo matching performance compared to the traditional methods. Hence, the expectation is that the aforementioned algorithms will increase the accuracy of the estimations of pine crown architecture traits. It may also be beneficial to use artificial lighting to capture images with more details that are vital for stereo matching. Furthermore, the camera could be raised vertically using a telescopic tripod, which would allow consistently high-resolution images to be captured at different tree heights at the expense of data collection time. One limitation of our system was that the stem and branches detected from the four viewing points around a tree were not registered in 3D space. The same branches may be visible in more than one stereo image pair. Future research could explore instance segmentation-assisted 3D reconstruction of pine tree architecture.

3.7 CONCLUSION

Our study developed and evaluated a high-throughput, non-contact method for measuring stem diameter, branch angle, and branch diameter of individual 6-year-old loblolly pine trees. The pipeline included the use of a stereo machine vision system in conjunction with a fine-tuned instance segmentation model and image-point-cloud processing techniques. The algorithm developed in this study yielded accurate and precise results regarding stem diameter estimation. The branch angle estimation achieved reasonable accuracy for objective classification into categories such as small, medium, and large branch angle. Occlusion, imaging angle, and limited spatial image resolution prevented accurate estimation of branch diameters for practical applications. In loblolly pine tree breeding trials, the proposed method may become a low-cost, objective, high-throughput tool for phenotyping tree architecture traits.

3.8 ACKNOWLEDGEMENTS

The authors are grateful to IFCO Seedlings, a member of the North Carolina State University Cooperative Tree Improvement Program, for making their loblolly pine progeny test site available for this study. We are thankful for the support from Alabama Experiment Station.

3.9 REFERENCES

Allen, H. L., Fox, T. R., and Campbell, R. G. (2005). What is ahead for intensive pine plantation silviculture in the south? *Southern Journal of Applied Forestry*, *29*(2), 62–69. https://doi.org/10.1093/SJAF/29.2.62

Apelt, F., Breuer, D., Nikoloski, Z., Stitt, M., and Kragler, F. (2015). Phytotyping4D: a light-field imaging system for non-invasive and accurate monitoring of spatio-temporal plant growth. *The Plant Journal*, *82*(4), 693–706. https://doi.org/10.1111/TPJ.12833

Bao, Y., Tang, L., Srinivasan, S., and Schnable, P. S. (2019). Field-based architectural traits characterisation of maize plant using time-of-flight 3D imaging. *Biosystems Engineering*, *178*, 86–101. https://doi.org/10.1016/J.BIOSYSTEMSENG.2018.11.005

Baweja, H. S., Parhar, T., Mirbod, O., and Nuske, S. (2018). StalkNet: A deep learning pipeline for high-throughput measurement of plant stalk count and stalk width. *Springer Proceedings in Advanced Robotics*, *5*, 271–284. https://doi.org/10.1007/978-3-319-67361-5_18/COVER

Bland, J. M., and Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, *8*(2), 135–160. https://doi.org/10.1177/096228029900800204

Borders, B. E., and Bailey, R. L. (2001). Loblolly Pine—Pushing the Limits of Growth. *Southern Journal of Applied Forestry*, *25*(2), 69–74. https://doi.org/10.1093/SJAF/25.2.69

Bosch, S., Serra Bragança, F., Marin-Perianu, M., Marin-Perianu, R., van der Zwaag, B. J., Voskamp, J., Back, W., van Weeren, R., and Havinga, P. (2018). EquiMoves: A wireless networked inertial measurement system for objective examination of horse gait. *Sensors 2018, Vol. 18, Page 850*, *18*(3), 850. https://doi.org/10.3390/S18030850

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

Brooks, J. (2019). *COCO Annotator*.

Chen, B., and Jung, C. (2018). Patch-based stereo matching using 3d convolutional neural networks. *Proceedings - International Conference on Image Processing, ICIP*, 3633–3637. https://doi.org/10.1109/ICIP.2018.8451527

Chen, C., Liu, M. Y., Tuzel, O., and Xiao, J. (2017). R-CNN for small object detection. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10115 LNCS*, 214–230. https://doi.org/10.1007/978-3-319-54193-8_14/COVER

Chu, P., Li, Z., Lammers, K., Lu, R., and Liu, X. (2021). Deep learning-based apple detection using a suppression mask R-CNN. *Pattern Recognition Letters*, *147*, 206–211. https://doi.org/10.1016/J.PATREC.2021.04.022

Cumbie, W.P., Isik, F. and McKeand, S.E., 2012. Genetic improvement of sawtimber potential in loblolly pine. Forest Science, 58(2), pp.168-177. https://doi.org/10.5849/forsci.09-060

Dai, H., Zhang, X., Zhao, Y., Sun, H., and Zheng, N. (2022). Adaptive disparity candidates prediction network for efficient real-time stereo matching. *IEEE Transactions on Circuits and Systems for Video Technology*, *32*(5), 3099–3110. https://doi.org/10.1109/TCSVT.2021.3102109

de Melo, R. F., de Lima, G. L., Corrêa, G. R., Zatt, B., de Aguiar, M. S., Nachtigall, G. R., and Araújo, R. M. (2020). Diagnosis of apple fruit diseases in the wild with mask r-cnn. *Lecture Notes in*

Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), *12319 LNAI*, 256–270. https://doi.org/10.1007/978-3-030-61377-8_18/COVER

Ehrenberg, C. 1970. Breeding for stem quality. Unasylva 24 (2-3), 23–31

Fischler, M. A., and Bolles, R. C. (1981). Random sample consensus. *Communications of the ACM*, *24*(6), 381–395. https://doi.org/10.1145/358669.358692

Fox, T.R., Jokela, E.J. and Allen, H.L., 2007. The development of pine plantation silviculture in the southern United States. Journal of Forestry, 105(7), pp.337-347. https://doi.org/10.1093/jof/105.7.337

Ganesh, P., Volle, K., Burks, T. F., and Mehta, S. S. (2019). Deep Orange: Mask R-CNN based orange detection and segmentation. *IFAC-PapersOnLine*, *52*(30), 70–75. https://doi.org/10.1016/J.IFACOL.2019.12.499

Gibbs, J. A., Pound, M., French, A. P., Wells, D. M., Murchie, E., and Pridmore, T. (2018). Plant phenotyping: An active vision cell for three-dimensional plant shoot reconstruction. *Plant Physiology*, *178*(2), 524–534. https://doi.org/10.1104/PP.18.00664

Gorte, B., and Pfeifer, N., 2004. Structuring laser-scanned trees using mathematical morphology. *In Proc. 20th ISPRS Congress, Istanbul, Turkey*, pp. 929–933. ISPRS.

Hatrisse, C., Macaire, C., Sapone, M., Hebert, C., Hanne-Poujade, S., de Azevedo, E., Marin, F., Martin, P., and Chateau, H. (2022). Stance phase detection by inertial measurement unit placed on the metacarpus of horses trotting on hard and soft straight lines and circles. *Sensors*, *22*(3), 703. https://doi.org/10.3390/S22030703/S1

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(2), 386–397. https://doi.org/10.48550/arxiv.1703.06870

Hirschmuller, H. (2008). Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(2), 328–341. https://doi.org/10.1109/TPAMI.2007.1166

Kaminuma, E., Heida, N., Tsumoto, Y., Yamamoto, N., Goto, N., Okamoto, N., Konagaya, A., Matsui, M., and Toyoda, T. (2004). Automatic quantification of morphological traits via three-dimensional measurement of Arabidopsis. *The Plant Journal*, *38*(2), 358–365. https://doi.org/10.1111/J.1365-313X.2004.02042.X

Kankare, V., Joensuu, M., Vauhkonen, J., Holopainen, M., Tanhuanpää, T., Vastaranta, M., Hyyppä, J., Hyyppä, H., Alho, P., Rikala, J., and Sipi, M. (2014). Estimation of the Timber Quality of Scots Pine with Terrestrial Laser Scanning. *Forests*, 5(8), 1879–1895. https://doi.org/10.3390/f5081879

Klodt, M., and Cremers, D. (2015). High-resolution plant shape measurements from multi-view stereo reconstruction. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8928*, 174–184. https://doi.org/10.1007/978-3-319-16220-1_13/COVER

Kukkonen, M., Maltamo, M., Korhonen, L., and Packalen, P. (2022). Evaluation of UAS LiDAR data for tree segmentation and diameter estimation in boreal forests using stem-and crown-based methods. *Canadian Journal of Forest Research*, *52*(5), 674–684. https://doi.org/10.1139/CJFR-2021-0217/ASSET/IMAGES/CJFR-2021-0217TAB6.GIF

Li, J., and Tang, L. (2017). Developing a low-cost 3D plant morphological traits characterization system. *Computers and Electronics in Agriculture*, *143*, 1–13. https://doi.org/10.1016/J.COMPAG.2017.09.025

Li, J., Wang, P., Xiong, P., Cai, T., Yan, Z., Yang, L., Liu, J., Fan, H., and Liu, S. (2022). Practical stereo matching via cascaded recurrent network with adaptive correlation. https://doi.org/10.48550/ARXIV.2203.11483

Liang, X., Kankare, V., Hyyppä, J., Wang, Y., Kukko, A., Haggrén, H., Yu, X., Kaartinen, H., Jaakkola, A., Guan, F., Holopainen, M., and Vastaranta, M. (2016). Terrestrial laser scanning in forest inventories. *ISPRS Journal of Photogrammetry and Remote Sensing*, *115*, 63–77. https://doi.org/10.1016/J.ISPRSJPRS.2016.01.006

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8693 LNCS* (PART 5). https://doi.org/10.1007/978-3-319-10602-1_48

Liu, Y., Huang, W., Wang Yan Liu, Y., Wang, Y., and Huang Yuheng Wang, W. (2018). A new stereo matching method for RAW image data based on improved SGBM. *Https://Doi.Org/10.1117/12.2505701*, *10845*, 208–214. https://doi.org/10.1117/12.2505701

Liu, Y., Yang, G., Huang, Y., and Yin, Y. (2021). SE-Mask R-CNN. *Journal of Intelligent and Fuzzy Systems*, *41*(6), 6715–6725. https://doi.org/10.3233/JIFS-210597

Lou, L., Liu, Y., Shen, M., Han, J., Corke, F., and Doonan, J. H. (2015). Estimation of branch angle from 3d point cloud of plants. *Proceedings - 2015 International Conference on 3D Vision, 3DV 2015*, 554–561. https://doi.org/10.1109/3DV.2015.68

Malhi, Y., Jackson, T., Bentley, L. P., Lau, A., Shenkin, A., Herold, M., Calders, K., Bartholomeus, H., and Disney, M. I. (2018). New perspectives on the ecology of tree structure and tree communities through terrestrial laser scanning. *Interface Focus*, *8*(2). https://doi.org/10.1098/RSFS.2017.0052

McKeand, S.E., 2019. The evolution of a seedling market for genetically improved loblolly pine in the southern United States. Journal of Forestry, 117(3), pp.293-301. https://doi.org/10.1093/jofore/fvz006

Nguyen, N. D., Do, T., Ngo, T. D., and Le, D. D. (2020). An evaluation of deep learning methods for small object detection. *Journal of Electrical and Computer Engineering*, *2020*. https://doi.org/10.1155/2020/3189691

Paul, A., Ghosh, S., Das, A. K., Goswami, S., das Choudhury, S., and Sen, S. (2020). A review on agricultural advancement based on computer vision and machine learning. *Advances in Intelligent Systems and Computing*, *937*, 567–581. https://doi.org/10.1007/978-981-13-7403-6_50/COVER

Paulus, S., Behmann, J., Mahlein, A. K., Plümer, L., and Kuhlmann, H. (2014). Low-cost 3d systems: suitable tools for plant phenotyping. *Sensors 2014, Vol. 14, Pages 3001-3018*, *14*(2), 3001–3018. https://doi.org/10.3390/S140203001

Peuhkurinen, J., Maltamo, M., Malinen, J., Pitkänen, J., and Packalén, P. (2007). Preharvest measurement of marked stands using airborne laser scanning. *Forest Science*, *53*(6). https://doi.org/10.1093/forestscience/53.6.653

Pfeifer, N., Gorte, B., and Winterhalder, D. (2004). Automatic reconstruction of single trees from terrestrial laser scanner data. *In Proceedings of 20th ISPRS Congress (Vol. 35, pp. 114-119). Istanbul: ISPRS.*

Radosevic, G., and Rizvić, S. (2012). *Laser scanning versus photogrammetry combined with manual post-modeling in stecak digitization. 14th Central European Seminar on Computer Graphics, Proceedings*.

Ranftl, R., Gehrig, S., Pock, T., and Bischof, H. (2012). Pushing the limits of stereo using variational stereo estimation. *IEEE Intelligent Vehicles Symposium, Proceedings*, 401–407. https://doi.org/10.1109/IVS.2012.6232171

Raumonen, P., Kaasalainen, M., Markku, Å., Kaasalainen, S., Kaartinen, H., Vastaranta, M., Holopainen, M., Disney, M., and Lewis, P. (2013). Fast automatic precision tree models from terrestrial laser scanner data. *Remote Sensing 2013, Vol. 5, Pages 491-520*, *5*(2), 491–520. https://doi.org/10.3390/RS5020491

Salas Fernandez, M. G., Bao, Y., Tang, L., and Schnable, P. S. (2017). A high-throughput, field-based phenotyping technology for tall biomass crops. *Plant Physiology*, *174*(4), 2008–2022. https://doi.org/10.1104/PP.17.00707

Schomaker, M. E., Zarnoch, S. J., Bechtold, W. A., Latelle, D. J., Burkman, W. G., and Cox, S. M. (2007). Crown-condition classification: A guide to data collection and analysis (SRS-GTR-102; p. SRS-GTR-102). U.S. Department of Agriculture, Forest Service, Southern Research Station. https://doi.org/10.2737/SRS-GTR-102

Schultz, R. P. (1999). Loblolly - The pine for the twenty-first century. *New Forests*, *17*(1–3). https://doi.org/10.1023/A:1006533212151

Sodhi, P., Vijayarangan, S., and Wettergreen, D. (2017). In-field segmentation and identification of plant structures using 3D imaging. *IEEE International Conference on Intelligent Robots and Systems*, *2017-September*, 5180–5187. https://doi.org/10.1109/IROS.2017.8206407

Southern Pine Inspection Bureau (SPIB). 1994. Grading rules. SPIB, Pensacola, FL.

Srinivasan, S., Popescu, S. C., Eriksson, M., Sheridan, R. D., and Ku, N. W. (2014). Multi-temporal terrestrial laser scanning for modeling tree biomass change. *Forest Ecology and Management*, *318*, 304–317. https://doi.org/10.1016/J.FORECO.2014.01.038

Steele, P.H. 1984. Factors determining lumber recovery in sawmilling. Gen. Tech. Rept. FPL-39. USDA Forest Serv., Forest Products Lab., Madison, WI. 8 pp. https://doi.org/10.2737/FPL-GTR-39

Sun, X., Fang, W., Gao, C., Fu, L., Majeed, Y., Liu, X., Gao, F., Yang, R., and Li, R. (2022). Remote estimation of grafted apple tree stem diameter in modern orchard with RGB and point cloud based on SOLOv2. *Computers and Electronics in Agriculture*, *199*, 107209. https://doi.org/10.1016/J.COMPAG.2022.107209

Tienaho, N., Yrttimaa, T., Kankare, V., Vastaranta, M., Luoma, V., Honkavaara, E., Koivumäki, N., Huuskonen, S., Hynynen, J., Holopainen, M., Hyyppä, J., and Saarinen, N. (2022). Assessing structural complexity of individual Scots pine trees by comparing terrestrial laser scanning and photogrammetric point clouds. *Forests 2022, Vol. 13, Page 1305*, *13*(8), 1305. https://doi.org/10.3390/F13081305

Triki, A., Bouaziz, B., Gaikwad, J., and Mahdi, W. (2021). Deep leaf: Mask R-CNN based leaf detection and segmentation from digitized herbarium specimen images. *Pattern Recognition Letters*, *150*, 76–83. https://doi.org/10.1016/J.PATREC.2021.07.003

Trincado, G., and Burkhart, H. E. (2008). A model of knot shape and volume in Loblolly pine trees. *Wood and Fiber Science*, 634–646. https://wfs.swst.org/index.php/wfs/article/view/2070

Nguyen, T. T., Slaughter, D. C., Maloof, J. N., and Sinha, N. (2016). Plant phenotyping using multi-view stereo vision with structured lights. *Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping*, *9866*(17), 22–30. https://doi.org/10.1117/12.2229513

Uusitalo, J. (1997). Pre-harvest measurement of pine stands for sawing production planning. *Acta Forestalia Fennica*, *259*. https://doi.org/10.14214/aff.7519

Uusitalo, J., and Isotalo, J. (2005). Predicting knottiness of Pinus sylvestris for use in tree bucking procedures. *Scandinavian Journal of Forest Research*, *20*(6). https://doi.org/10.1080/02827580500407109

Wang, Y., Chao, W. L., Garg, Di., Hariharan, B., Campbell, M., and Weinberger, K. Q. (2019). Pseudo-lidar from visual depth estimation: Bridging the gap in 3D object detection for autonomous driving. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2019-June*, 8437–8445. https://doi.org/10.1109/CVPR.2019.00864

Warensjö, M., and Rune, G. (2004). Stem straightness and compression wood in a 22-year-old stand of container-grown Scots pine trees. *Silva Fennica*, *38*(2), 143–153. https://doi.org/10.14214/SF.424

Wheeler, N. C., Steiner, K. C., Schlarbaum, S. E., and Neale, D. B. (2015). The evolution of forest genetics and tree improvement research in the United States. *Journal of Forestry*, *113*(5), 500–510. https://doi.org/10.5849/JOF.14-120

White, T., Davis, J., Gezan, S., Hulcr, J., Jokela, E., Kirst, M., Martin, T. A., Peter, G., Powell, G., Smith, J., White, T., Davis, J., Gezan, Á. S., Hulcr, Á. J., Jokela, Á. E., Kirst, Á. M., Martin, T. A., Peter, Á. G., Powell, Á. G., and Smith, Á. J. (2013). Breeding for value in a changing world: past achievements and future prospects. *New Forests 2013 45:3*, *45*(3), 301–309. https://doi.org/10.1007/S11056-013-9400-X

Wilkes, P., Lau, A., Disney, M., Calders, K., Burt, A., Gonzalez de Tanago, J., Bartholomeus, H., Brede, B., and Herold, M. (2017). Data acquisition considerations for Terrestrial Laser Scanning of forest plots. *Remote Sensing of Environment*, *196*, 140–153. https://doi.org/10.1016/J.RSE.2017.04.030

Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., and Girshick, R. (2019). *Detectron2*.

Xiang, L., Tang, L., Gai, J., and Wang, L. (2021). Measuring stem diameter of sorghum plants in the field using a high-throughput stereo vision system. *Transactions of the ASABE*, *64*(6), 1999–2010. https://doi.org/10.13031/TRANS.14156

Zhang, J., He, L., Karkee, M., Zhang, Q., Zhang, X., and Gao, Z. (2017). Branch detection with apple trees trained in fruiting wall architecture using stereo vision and regions-convolutional neural network(R-CNN). *2017 ASABE Annual International Meeting*, 1-. https://doi.org/10.13031/AIM.201700427

Ziamtsov, I., and Navlakha, S. (2020). Plant 3D (P3D): A plant phenotyping toolkit for 3D point clouds. *Bioinformatics*, *36*(12), 3949–3950. https://doi.org/10.1093/BIOINFORMATICS/BTAA220

Zobel, B.J., and J.T. Talbert. 1984. Applied tree improvement. John Wiley and Sons, New York. 505 p. https://doi.org/10.1016/0378-1127(85)90086-6

CHAPTER 4. GENERAL CONCLUSION

4.1 CONCLUSIONS

Using 3D stereo videography and a processing pipeline, automated EKGA technology was developed. A combination of several modules was used to build this EKGA system, including a DLC model to identify body landmarks, a Faster R-CNN model to classify hoof cycle phases, and a SGBM model to compute 3D coordinates of the detected landmarks. To demonstrate the proposed methodology's effectiveness, we used the pipeline's outputs SL and SD as case studies, and we applied post-processing techniques to enhance their performance. Head length measurements in the field confirmed that the 2D detection algorithm combined with stereo matching performed satisfactorily and consistently. This pipeline did a precise job of classifying the hooves. It is evident from a B-A study that the combination of modules suggested in this paper could potentially be used for evaluating horse kinematic parameters, with a bias of 0.11 s for SD and -0.014 m for SL.

Our pipeline developed for this study can be useful for research in gait genetics. Identifying trends in populations or subpopulations requires precise phenotyping in equine genetic studies that examine locomotion, such as gait. Ultimately, the proposed method is a low-cost, practical, and rapid tool that animal scientists can use to conduct EKGAs. In spite of the fact that this system might sometimes be less precise than marker-based optical motion capture systems available on the market, it requires less effort to use. Gait phenotyping is still mostly subjective in most studies, making it less accurate than desired. In this way, we believe we will be able to gain greater insight into how genetics relates to gait with our more accurate methods.

Using a high-throughput approach, we developed and evaluated a non-contact way to measure the stem diameter, branch angle, and branch diameter of an individual 4-year-old loblolly pine tree. Using stereo machine vision, a fine-tuned instance segmentation model, and image/point cloud processing tools were used as part of this pipeline. In terms of stem diameter estimation, our proposed algorithm produced accurate and precise results. For the classification of branches into low, medium, and high levels, the branch angle estimation achieved reasonable accuracy. Occlusion, imaging angle, and limited spatial image resolution prevented the accurate estimation of branch diameters for practical applications. Loblolly pine tree breeding trials could benefit from a low-cost, high-throughput phenotyping tool for tree architecture traits based on the proposed method. Using this system in progeny trials will make it easier to select genotypes exhibiting desirable tree architectures because the system provides safe, rapid, and quantitative information.

4.2 RECOMMENDATIONS

Additional assessment algorithms and data processing methods, including a combination of multiple cameras, are needed to improve the accuracy of kinematic gait parameters, notably SL and SD measurements. Multiple wirelessly synchronized stereo cameras can be added to the stereo imaging pipeline. It is possible to track and analyze landmarks around a horse's body using this method. To improve stereo matching accuracy and disparity maps quality, a more precise stereo matching algorithm can also be utilized. Based on the computational system used, SGBM was preferred due to its high efficiency. To improve the accuracy of stereo image similarity computation, state-of-the-art CNNs can be used. It may also not be necessary to use dense stereo matching as landmarks' 3D coordinates are the key to the process, and their disparity might be determined directly. Aside from this, the distance between the camera and the horse determines the spatial resolution (i.e., the level of detail) and the accuracy of stereo matching-based depth estimation, so a study of the camera-horse distance in relation to the proposed system is of interest as well.

The development of our developed system could lead to an observation of the stem's structure using a derived 3D point cloud, thus allowing the stem's straightness to be determined. As a conifer stem is tilted from its normal vertical orientation, compression wood forms on the lower side, inhibiting the growth of the terminal shoot. Warensjö and Rune (2004) state that a serious defect called compression wood causes downgrading. It is therefore crucial to evaluate stem straightness. There may be some further improvements that can be made to the proposed pipeline in order to increase its precision and accuracy. Recent research has developed a few thin object stereo matching algorithms. A typical stereo algorithm (SGBM) contains 18.4% of bad pixels in scenes with reflective objects, compared to 9.27% in Zhou et al. (2017). The stereo matching algorithms have also been implemented on RAW images to improve binocular stereo vision system accuracy (Yan Liu et al., 2018). The performance of stereo matching was also enhanced by Chen and Jung (2018) using convolutional neural networks (CNNs). It is speculated that integrating the aforementioned algorithms can improve the accuracy of estimations of pine tree traits. The use of artificial lighting may also help capture images with more details, which are essential for stereo matching. A telescopic tripod could also be used to transport the camera vertically, allowing high-resolution images to be captured at different tree heights at the expense of data collection time. Due to the fact that our system collected images 90 degrees apart around the trees, we were unable to track the branches. An adequate overlap between images is required if a data collection system is to track branches. As more images are incorporated into the measurement pipeline, structure from motion and instance segmentation can be combined to improve accuracy in estimating branch angles.

VITA

Nariman Niknejad is a M.Sc. Student in AgCypherLab located at the department of Biosystems Engineering of Auburn University, Alabama, USA. He received his B.Sc. degree in Mechanical Engineering from K. N. Toosi University of Technology, Tehran, Iran. His research interests include Precision Agriculture, Image Processing, Deep Learning, and Mechatronics.

List of Conference Papers:

1. Niknejad, N., Caro, J., Puhl, R. B., Bao, Y., and Staiger, E. A. (2022). Estimation of Equine Stride Length and Stance Duration Using Stereo 3D Videography and Deep Learning. 2022 ASABE Annual International Meeting, 1. American Society of Agricultural and Biological Engineers.
2. Akter, M., Niknejad, N., Bao, Y., Puhl, R. B., Payn, K., and Zheng, J. (2021). Phenotyping of Pine Tree Architecture with Stereo Vision and Deep Learning. 2021 ASABE Annual International Virtual Meeting, 1. American Society of Agricultural and Biological Engineers.