Precision Agriculture Systems for the Southeast US Using Computer Vision and Deep Learning

by

Rafael Bidese Puhl

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy
in Biosystems Engineering

Auburn, Alabama
August 5, 2023

Keywords: Deep Learning, Machine Vision, Precision Agriculture, Sensors

Approved by

Yin Bao, Chair, Assistant Professor of Biosystems Engineering at Auburn University
Jeremiah Davis, Co-Chair, Professor of Biosystems Engineering at Auburn University
Timothy McDonald, Professor of Biosystems Engineering at Auburn University
Tanzeel Rehman, Assistant Professor of Biosystems Engineering at Auburn University
Jingyi Zheng, Assistant Professor of Mathematics and Statistics at Auburn University
Joseph Purswell, Supervisory Agricultural Engineer of Poultry Research at USDA-ARS

Abstract

Recent advancements in sensors, machine vision and deep learning with the development of efficient algorithms has enabled new opportunities in precision agriculture. In this dissertation three different projects are presented.

First, an automated pine seedling counting for nursery management. In nursery management, accurate inventory of seedlings provides insights into how many seedlings can be sold and/or if there is any loss due to washout, mechanical damage or pest/diseases that can still be mitigated. In this study we developed a system to count pine seedlings in production sites and map the seedling density in the field. The mean absolute percentage error (MAPE) of our best performing model was 7.53%, which is an improvement over the baseline manual sampling-based approach with a MAPE of 11.07%. The results showed that the proposed approach was able to count seedlings in a crowded scene under complex field conditions with higher accuracy than the standard manual practice.

Second, an mmWave radar-based peanut yield monitor, an essential equipment for precision agriculture is presented. During the harvest of peanuts, it is common for foreign materials to run through the pneumatic conveyor in the harvester, making the environment inside the pneumatic conveyor harmful for components that are fragile. Therefore, a millimeter-wave FMCW radar-based mass-flow sensor was developed to monitor yield during peanut harvest. After evaluation of 5-fold cross validation, the best performing models achieved an RMSE of 0.14 kg/s (19 lb/min) and a sMAPE of 15% while having an R^2 of 0.85 for the research-scale combine and an RMSE of 0.52 kg/s (69 lb/min) and a sMAPE of 10% while having an R^2 of 0.71 for the commercial-scale combine. Though the results are promising, further field investigation is necessary to evaluate the effects of different field conditions.

Finally, a broiler activity index measurement is demonstrated using a detect-and-track pipeline and a age-compensated activity index. Precision management in poultry farming, utilizing advanced technologies like computer vision and lighting control, has emerged as a promising approach to enhance productivity, meat quality, and animal welfare. This work generated multi-animal detection-and-tracking dataset from top-view videos of broilers at different growth stages and different times of a day. Accuracy and efficiency of different state-of-the-art object detection (YOLOv8 and YOLO-NAS) and trackers (SORT and ByteTrack) were evaluated and compared regarding the generation of a new broiler activity index. The fastest pipeline (YOLOv8n+SORT) achieved 123 frames per second with a correlation of 96.4% with the proposed broiler activity index. Once the behaviors are characterized the pipeline can be used to generate feedback signal for control of the broiler houses, providing better animal welfare.

Acknowledgments

I wish to express my profound gratitude to the many individuals who have offered their guidance, assistance, and encouragement during my research journey. Your invaluable insights and feedback have been pivotal in shaping this dissertation.

Foremost, I am deeply grateful to my advisor, Dr. Yin Bao, for his unwavering support, guidance, and mentorship. His expertise, patience, and unwavering dedication have been instrumental in shaping my research trajectory and fostering my academic growth.

My heartfelt thanks go to my dissertation committee members, Jeremiah Davis, Timothy McDonald, Tanzeel Rehman, Jingyi Zheng, Joseph Purswell, and Alvaro Sans-Saez, for their valuable contributions of time, expertise, and constructive feedback. Their insightful suggestions have greatly enriched the quality of this work and contributed significantly to my scholarly development.

I am also grateful to the individuals and organizations with whom I collaborated during this research. I extend my thanks to USDA-ARS National Peanut Research Laboratory, Kelley Manufacturing Co., Southern Forest Nursery Management Cooperative, USDA-ARS Poultry Research Unit, and Biosystems Engineering Department of Auburn University, along with the respective individuals: Chris Butts, Joseph McIntyre, Phat Dang, Alvaro Sanz-Saez, Bennie Branch, Matt Rewis, Jason Morris, Nina Payne, Thomas Stokes, Scott Enebak, Jeremiah Davis, Kelly Griggs, Cody Smith, James Johnson, Bobby Bradford, and Caroline Whiting. Your collective efforts and shared expertise have profoundly enhanced the scope and impact of this dissertation.

My experience at Auburn University was enriched by several amazing individuals who contributed significantly to my personal, academic, and professional growth. My deepest gratitude goes to Dr. Nedret Billor and Roberto Molinari, who guided me professionally

throughout the MS in Data Science capstone project, and Dr. Rahul Sukthankar, who under his mentorship, instilled in me an unshakeable confidence in my capabilities.

A special acknowledgment goes to my labmates, Mary Beth Cassity, Kamand Bagherian, Nariman Niknejad, Sharif Shibani, Puranjit Singh, and Spencer Overton. Their technical assistance, support, camaraderie, and lively discussions enriched my research experience. Shared moments of field work, conference attendance, and even long drives were made more memorable and enjoyable through their companionship.

To my family, who provided unwavering love, understanding, and encouragement throughout my academic journey, I extend my deepest appreciation. Their relentless belief in my abilities has been my strongest motivation and source of inspiration. I especially thank my wife, Gabi, and my children, Marcelo and Felipe, for their enduring patience and support. I am also profoundly grateful to my mother, Cristina, and brother, Eduardo, back in Brazil, for their continuous encouragement over the years.

In closing, I want to reiterate my indebtedness to each and every individual named above. Your collective efforts, support, and faith in my abilities have played an integral part in the successful completion of this dissertation. Thank you.

Table of Contents

List of Figures

List of Tables

Chapter 1

GENERAL INTRODUCTION

1.1 Background

Recent developments in deep learning (DL) have the potential to revolutionize the landscape of precision agriculture and applied computer vision. DL techniques, particularly convolutional neural networks (CNNs), have demonstrated remarkable capabilities in image analysis and pattern recognition. In the context of precision agriculture, DL models can be trained to automatically analyze vast amounts of visual data collected from sensors, drones, or satellites, enabling efficient and accurate crop monitoring and management (e.g., automated architectural traits extraction, and organ detection and counting) (Li et al., 2020). For counting in static images, the main tasks can be formulated as regression, object detection, and semantic segmentation. Regression has been widely used for counting plants in high density, such as wheat (Lu and Cao, 2020), rice (Liu et al., 2020), and maize (Lu et al., 2017). For counting in videos, the image sequence adds another level of complexity of associating the same objects across consecutive frames. This problem is typically solved by using tracking algorithms or by reconstruction of the scene in three-dimensional (3D) space (Sun et al., 2020; Tan et al., 2022).

Within the sphere of precision poultry farming, the ability to monitor and track the movements and activities of birds can provide significant insights into their health, well-being, and efficiency (Guo et al., 2022). This is accomplished through the application of advanced technological tools such as machine learning, computer vision, and sensor systems (Li et al., 2022; Okinda et al., 2020; van der Sluis et al., 2020). Computer vision has been notably recognized as a potent tool for identifying and tracing birds within a poultry farming environment (Aydin, 2017; Del Valle et al., 2021; Okinda et al., 2020; Zhang et

al., 2022). By processing video data and interpreting it, computer vision algorithms can differentiate individual birds, trace their trajectories, and even decipher certain behaviors. This allows for a non-invasive, continuous supervision of bird behavior, contributing to the overall evaluation of flock health and early detection of any signs of distress or illness. For instance, alterations in patterns or levels of activity could signal health-related concerns. By monitoring these behaviors, timely intervention can be ensured, thereby enhancing animal welfare and productivity.

Sensor development in precision agriculture has seen advancements with the incorporation of modern technologies, such as radio detection and ranging (radar) systems. Longchamps et al. (2022) reviewed several contemporary approaches for yield estimation, including the utilization of millimeter-wave (mmWave) radar. This technology has been applied in non-contact proximal sensing for crops that grow underground, such as potatoes, manioc, and sugar beets. For instance, Konstantinovic et al. (2008) employed ultra-wideband (UWB) radar to detect the location of sugar beets, achieving significant correlation between beet biomass and backscattered energy. The use of machine vision techniques for radar signal processing was suggested for further enhancement. Radar sensors have also found application in assessing plant vigor in vineyards. Henry et al. (2019) demonstrated a radar-based system that scanned grape canopies to estimate biomass. The mmWave radar approach showed promise by detecting occluded grapes and being minimally affected by dust particles due to its small wavelength. Moreover, the radar sensor's robust, dust-proof enclosure enables its resilience in the challenging harvesting environment. These developments highlight the potential of mmWave radar as a valuable tool in sensor development for precision agriculture.

1.2 Research Objectives

1.2.1 Computer Vision for Automated Pine Seedling Inventory

Pine plantations cover 16 million hectares (84%) of the total planted forests across the Southern United States (Miles, 2019). To meet the demand for reforestation, the Southeastern U.S. produces over 1 billion seedlings per year (Fargione et al., 2021). The state of Alabama alone is responsible for 91 million seedlings, ranking as the 4^{th} largest producer in the U.S. (Haase et al., 2020). Recent efforts in improving seedling genetics have increased the value of individual seedlings, ranging from \$50 to \$435 per thousand seedlings depending on the pollination process used to guarantee the improved genetics (e.g., open pollination, \$50-\$75; mass control pollination, \$120-\$140; and somatic embryogenesis, up to \$430) (Rousseau et al., 2012). With an increase in crop values in the field, it is paramount for forest tree nurseries to obtain accurate seedling inventories. It is a common practice in forest nurseries to gather seedling inventories twice per year (in spring and fall) to determine the number of seedlings that are available and can potentially be sold. The spring seedling inventory is necessary for nursery managers to determine seedling bed densities and update seedling losses due to washout, mechanical damage, and pest/disease issues. This inventory is the first assessment of crop development and aids managers in determining areas of concern and where mitigation can still occur. The fall inventory is more important than the inventory in spring, as it is needed to estimate the number of shippable seedlings and anticipate final sales.

The widely used seedling counting practice today was established more than half a century ago (Wakeley, 1954). It requires selecting multiple 0.3 m $\times$ 1.22 m (1- by 4-foot) plots for each seed lot used at the nursery. For each sample plot, the total number of seedlings is manually counted. A worker must bend over, sort through densely populated seedlings, and count the narrow stems that are growing within the counting frame placed on the bed. The inventory task is labor-intensive, time-consuming, error-prone, and ergonomically

poor. Commercial nurseries have hundreds of kilometers of beds to cover with a limited staff. Therefore, automated seedling inventory technology can significantly reduce labor costs and improve working conditions for forest tree nursery production. The first forest tree seedling counter was developed by the USDA Forest Services Missoula Technology and Development Center (Gasvoda and Herzberg, 1993). A side-viewing infrared optical sensor was used to detect seedling stems. The machine counting was within $\pm 10\%$ of the manual counting. However, due to the side-viewing mechanism, large drill row spacing (i.e., 0.152 m) was required to carry the optical sensor. This system cannot operate on typical modern commercial seedling beds where there is no space to position side viewing sensor. To the best of our knowledge, there is no other related work in the literature.

The U.S. bareroot pine seedling production system employs a high planting density (e.g., 269 seedlings per square meter). In fall, when the seedling canopy covers the entire seedling bed, it is impossible for humans to visually separate individual seedlings from a top-view while standing on the ground in a non-contact manner, let alone for unmanned aerial vehicles with a high-resolution camera. We employed a push bar that bends the seedlings downward to expose the individual seedling stems. As the push bar moves forward, seedlings are released and return to their upright position. The movements of those seedlings are video-taped and used as information to perform the seedling count. A typical forest nursery can have 1.6 million seedlings per ha, making counting every seedling impossible. In this work, contrary to other counting systems that employ the detect-and-track approach, we present a machine vision system to count pine seedlings using end-to-end deep learning regression based on a CNN coupled with LSTM. The specific research objectives were to: (1) develop a deep learning-based machine vision system to count mechanically pushed pine seedlings, (2) simulate the accuracy of the standard manual sampling-based seedling counting, and (3) evaluate the machine counting performance against manual ground truth counting and the standard method.

1.2.2 mmWave Radar-Based Sensor Development Towards Peanut Yield Monitor

A yield monitor mounted on a combine, together with a global positioning system (GPS) allows the generation of yield maps that are used for identification of crop performance issues in the field that can be further investigated and possibly mitigated through management practices. Yield monitors for major crops such as corn, soybean, and cotton have already been widely commercialized, but many crops still do not have a reliable system to monitor harvest. Economic relevance is a major factor for why certain crops to have more developed technologies. Field conditions during harvest make the development of sensors for certain crop production systems even more challenging. During the harvest of peanuts, it is common for foreign materials (e.g. soil, vines, rocks, etc.) to run through the pneumatic conveyor in the combine. This makes the environment inside the air duct of the pneumatic conveyor harmful for components that are fragile. For instance, optical systems that have lenses or sensors inside the duct must employ strategies to reduce wearing of the parts due to the abrasive nature of the environment, but they are not enough to withstand this harsh environment for long periods of time. Other load cells-based systems have been tested in the field but provide low spatial resolution, making management practices less accurate. Some of the peanut yield monitors developed employed load cells mounted to the combine basket.

Researchers have developed various yield monitoring systems for agriculture. Vellidis et al. (2001) designed a production-scale system with errors around 5% for loads ranging from 1,567 to 5,420 kg. However, the resulting yield maps were noisy due to combine movement. Kirk et al. (2012) focused on small research plots but had low spatial resolution. Optical-based systems, such as those developed by Rains et al. (2005), Thomasson et al. (2006), Porter et al. (2017), and Porter et al. (2020), required field-specific calibration and faced challenges with lens damage, dirt buildup, and the harsh environment of the pneumatic conveyor. Free et al. (2014) adapted an impact plate-based system and compensated for airspeed variations, achieving moderate accuracy. These studies emphasize the need for

calibration, addressing environmental challenges, and improving accuracy in yield monitoring systems.

Currently, there is no commercially available peanut yield monitor from peanut combine manufacturers and agricultural technology companies. To comply with such environment, a non contact measurement of the yield is of need. Given the success of radar sensors in object detection (Henry et al., 2019; Konstantinovic et al., 2008), we propose to use a frequency modulated continuous wave (FMCW) millimeter-wave (mmWave) radar to measure mass flow rate of pneumatically conveyed unshelled peanuts. A data-driven approach was investigated to model the relationship between the radar image signature to a given mass flow rate. To handle the high dimensional radar range-velocity image, machine learning models are employed to perform the modeling task. The specific research objectives were to: (1) collect mmWave radar data from pneumatically conveyed peanuts at different mass flow rates, and (2) evaluate the performance of multiple machine learning models for prediction of peanut mass flow rate from the radar data.

1.2.3 Broiler Activity Index Based On Machine Vision And Detect-And-Track

In the field of precision poultry farming, the tracking and identification of birds' movements and behaviours can offer crucial insights into their health, welfare, and productivity (Guo et al., 2022). These insights are acquired through the deployment of advanced technologies such as computer vision, machine learning, and sensor-based systems (Li et al., 2022; Okinda et al., 2020; van der Sluis et al., 2020). Specifically, computer vision has established itself as a potent tool for detecting and tracking birds within a poultry farm environment (Aydin, 2017; Del Valle et al., 2021; Okinda et al., 2020; Zhang et al., 2022). By processing and analyzing video data, computer vision algorithms can identify individual birds, trace their movements over time, and even discern specific behaviours. This method enables the continuous, non-invasive monitoring of bird activity, which can be employed to evaluate overall flock health and detect any early signs of disease or distress. For example, alterations

in activity levels or patterns can be symptomatic of health issues. A sudden reduction in activity might indicate that a bird is unwell, while increased aggression could signal stress or overcrowding (Fan et al., 2022; Zukiwsky et al., 2020). By tracking these behaviours, farmers can intervene as required to address these issues, enhancing animal welfare and productivity.

Despite these advances, the influence of ambient light control as a modulator of activity in poultry farming is a relatively new research area. Light intensity has been demonstrated to impact the behaviour and physiological responses of poultry, which can, in turn, affect meat quality. Kuttappan et al., (2016) found a correlation between increased growth rate in birds and the onset of white striping and woody breast myopathies. This suggests that management practices that regulate growth rate, such as activity modulation via light control, could potentially mitigate these conditions. The woody breast condition, characterized by a tough consistency in raw breast fillets, has been a significant concern in the poultry industry due to its adverse effects on meat quality. Recent studies have begun to explore the potential of light control in managing this condition (Mirella Fernandes et al., 2021). For instance, Greene et al., (2019) discovered that the severity of woody breast myopathy could be reduced through the modulation of oxygen homeostasis-related genes in broiler chickens, suggesting a potential connection between light intensity, activity levels, and the development of woody breast.

To address these issues, this study presents a novel multi-animal detection and tracking pipeline specifically tailored for bird activity monitoring. A comprehensive dataset comprising top-view videos of broilers at different growth stages and various times of the day was generated. This dataset serves as the foundation for evaluating the accuracy and efficiency of state-of-the-art object detection models, including YOLOv8 (Jocher et al., 2023) and YOLO-NAS (Aharon et al., 2021), as well as tracking algorithms such as SORT (Bewley et al., 2016) and ByteTrack (Zhang et al., 2022).

The specific research objectives were to: (1) generate a comprehensive dataset from top-view videos of broilers at different growth stages and times of the day to create a multi-animal detection-and-tracking dataset for machine vision algorithm training and evaluation; (2) assess the performance of object detection models (e.g. YOLOv8 and YOLO-NAS) to identify the most accurate and efficient method for detecting individual broilers in complex scenes; (3) evaluate the accuracy and efficiency of SORT and ByteTrack algorithms for tracking multiple broilers simultaneously, identifying the most suitable algorithm for reliable broiler tracking;

1.3 Dissertation Outline

This dissertation comprises a collection of three journal articles that showcase our research in three distinct areas within the field of precision agriculture: (1) computer vision techniques applied to automate pine seedling inventory, (2) the advancement of mmWave Radar-based sensor for peanut yield monitoring, and (3) the utilization of machine vision and detect-and-track algorithms to create a broiler activity index. Chapter 1 serves as an introduction, outlining the research objectives. The first article (Chapter 2) focuses on the application of computer vision for automated pine seedling inventory. The second article (Chapter 3) presents the development of mmWave Radar-based sensor for peanut yield monitoring. Lastly, the third article (Chapter 4) describes the design and evaluation of a broiler activity index utilizing machine vision and detect-and-track algorithms.

REFERENCES

Aharon, S., Louis-Dupont, Ofri Masad, Yurkova, K., Lotem Fridman, Lkdci, Khvedchenya, E., Rubin, R., Bagrov, N., Tymchenko, B., Keren, T., Zhilko, A., & Eran-Deci. (2021). Super-Gradients. In GitHub repository. GitHub. https://doi.org/10.5281/ZENODO.7789328

Aydin, A. (2017). Using 3D vision camera system to automatically assess the level of inactivity in broiler chickens. Computers and Electronics in Agriculture, 135, 4–10. https://doi.org/10.1016/J.COMPAG.2017.01.024

Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016). Simple online and realtime tracking. 2016 IEEE International Conference on Image Processing (ICIP), 3464–3468. https://doi.org/10.1109/ICIP.2016.7533003

Del Valle, J. E., Pereira, D. F., Mollo Neto, M., Gabriel Filho, L. R. A., & Salgado, D. D. A. (2021). Unrest index for estimating thermal comfort of poultry birds (Gallus gallus domesticus) using computer vision techniques. Biosystems Engineering, 206, 123–134. https://doi.org/10.1016/J.BIOSYSTEMSENG.2021.03.018

Fargione, J., Haase, D. L., Burney, O. T., Kildisheva, O. A., Edge, G., Cook-Patton, S. C.,... Guldin, R. W. (2021). Challenges to the reforestation pipeline in the United States. Front. For. Glob. Change, 4. https://doi.org/10.3389/ffgc.2021.629198

Free, D., Kirk, K. R., White, J. W., Brantley, S. A., Peele, J. S., Monfort, W. S., Thomas, J. S., Massey, H. F., Han, Y. J., Montreal, A., & Canada, Q. (2014). Testing of an Impact Plate Yield Monitor for Peanuts: Mounting Configurations and Air Pressure Correction Written for presentation at the 2014 ASABE and CSBE/SCGAB Annual International Meeting Sponsored by.

Gasvoda, D., & Herzberg, D. (1993). Seedling counter field test. Tree Planters' Notes, 44(1).

Guo, Y., Aggrey, S. E., Wang, P., Oladeinde, A., & Chai, L. (2022). Monitoring Behaviors of Broiler Chickens at Different Ages with Deep Learning. Animals 2022, Vol. 12, Page 3390, 12(23), 3390. https://doi.org/10.3390/ANI12233390

Haase, D. L., Pike, C., Enebak, S., Mackey, L., Ma, Z., & Silva, C. (2020). Forest nursery seedling production in the United States - Fiscal year 2019. Tree Planters' Notes, 63(2), 26-31. Retrieved from https://rngr.net/publications/tpn/63-2/forest-nursery-seedling-production-in-the-united-states2014fiscal-year-2019

Henry, D., Aubert, H., & Veronese, T. (2019). Proximal Radar Sensors for Precision Viticulture. IEEE Transactions on Geoscience and Remote Sensing, 57(7), 4624–4635. https://doi.org/10.1109/TGRS.2019.2891886

Jocher, G., Chaurasia, A., & Qiu, J. (2023). YOLO by Ultralytics. https://github.com/ultralytics/ultralytics

Kirk, K. R., Porter, W. M., Monfort, W. S., Han, Y. J., Henderson, W. G., & Thomas, J. (2012). Development of a Yield Monitor for Peanut Research Plots. American Society of Agricultural and Biological Engineers Annual International Meeting 2012, ASABE 2012, 3, 1-. https://doi.org/10.13031/2013.41823

Konstantinovic, M., Woeckel, S., Schulze Lammers, P., & Sachs, J. (2008). UWB Radar System for Yield Monitoring of Sugar Beet. Transactions of the ASABE, 51(2), 753–761. https://doi.org/10.13031/2013.24372

Kuttappan, V. A., Hargis, B. M., & Owens, C. M. (2016). White striping and woody breast myopathies in the modern poultry industry: a review. Poultry Science, 95(11), 2724–2733. https://doi.org/10.3382/PS/PEW216

Li, Z., Guo, R., Li, M., Chen, Y., & Li, G. (2020). A review of computer vision technologies for plant phenotyping. Comput. Electron. Agric., 176, 105672. https://doi.org/10.1016/j.compag.2020.105672

Liu, L., Lu, H., Li, Y., & Cao, Z. (2020). High-throughput rice density estimation from transplantation to tillering stages using deep networks. Plant Phenomics, 2020. https://doi.org/10.34133/2020/1375957

Longchamps, L., Tisseyre, B., Taylor, J., Sagoo, L., Momin, A., Fountas, S., Manfrini, L., Ampatzidis, Y., Schueller, J. K., & Khosla, R. (2022). Yield sensing technologies for perennial and annual horticultural crops: a review. Precision Agriculture. https://doi.org/10.1007/s11119-022-09906-2

Lu, H., & Cao, Z. (2020). TasselNetV2+: A fast implementation for high-throughput plant counting from high-resolution RGB imagery. Front. Plant Sci., 11. https://doi.org/10.3389/fpls.2020.541960

Lu, H., Cao, Z., Xiao, Y., Zhuang, B., & Shen, C. (2017). TasselNet: Counting maize tassels in the wild via local counts regression network. Plant Methods, 13, 79. https://doi.org/10.1186/s13007-017-0224-0

Miles, P. D. (2019). Forest Inventory EVALIDator web-application Version 1.6. 0.03. St. Paul, MN: USDA.

Mirella Fernandes, A., de Lucca Sartori, D., José de Oliveira Morais, F., Salgado, A., Florentino Pereira, D., Oliveira Morais, de, Analysis, D., Cambra-López, M., Jorge De Moura, D., & Zheng, W. (2021). Analysis of Cluster and Unrest Behaviors of Laying Hens Housed under Different Thermal Conditions and Light Wave Length. Animals 2021, Vol. 11, Page 2017, 11(7), 2017. https://doi.org/10.3390/ANI11072017

Okinda, C., Nyalala, I., Korohou, T., Okinda, C., Wang, J., Achieng, T., Wamalwa, P., Mang, T., & Shen, M. (2020). A review on computer vision systems in monitoring of poultry: A welfare perspective. Artificial Intelligence in Agriculture, 4, 184–208. https://doi.org/10.1016/J.AIIA.2020.09.002

Porter, E., Vellidis, G., Liakos, V., Porter, W., & Branch, B. (2017). An Optical Yield Monitor for Peanuts-Proof of Concept and Evaluation. https://doi.org/10.1017/S2040470017000061

Porter, W. M., Ward, J., Taylor, R. K., & Godsey, C. B. (2020). A Note on the Application of an AgLeader® Cotton Yield Monitor for Measuring Peanut Yield: An Investigation in Two US states. Peanut Science, 47(2), 115–122. https://doi.org/10.3146/PS19-16.1

Rains, G. C., Perry, C. D., & Vellidis, G. (2005). Adaptation and testing of the Agleader cotton yield sensor on a peanut combine. In Applied engineering in agriculture: Vol. v. v. 21.

Rousseau, R. J., Roberts, S. D., & Herrin, B. L. (2012). Comparison of advanced genetic loblolly pine planting stock. In J. R. Butnor (Ed.), Proceedings of the 16th Biennial Southern Silvicultural Research Conference. Technical Report SRS-156. (pp. 269-272). Asheville, NC: USDA Forest Service, Southern Research Station.

Sun, S., Li, C., Chee, P. W., Paterson, A. H., Jiang, Y., Xu, R.,... Shehzad, T. (2020). Three-dimensional photogrammetric mapping of cotton bolls in situ based on point cloud segmentation and clustering. ISPRS J. Photogramm. Remote Sens., 160, 195-207. https://doi.org/10.1016/j.isprsjprs.2019.12.011

Tan, C., Li, C., He, D., & Song, H. (2022). Towards real-time tracking and counting of seedlings with a one-stage detector and optical flow. Comput. Electron. Agric., 193, 106683. https://doi.org/10.1016/j.compag.2021.106683

Thomasson, J. A., Sui, R., Wright, G. C., & Robson, A. J. (2006). Optical Peanut Yield Monitor: Development and Testing. Applied Engineering in Agriculture, 22(6), 809–818.

van der Sluis, M., de Haas, Y., de Klerk, B., Bas Rodenburg, T., & Ellen, E. D. (2020). Assessing the Activity of Individual Group-Housed Broilers Throughout Life Using a Passive Radio Frequency Identification System—A Validation Study. Sensors 2020, Vol. 20, Page 3612, 20(13), 3612. https://doi.org/10.3390/S20133612

Vellidis, G., Perry, C. D., Durrence, J. S., Thomas, D. L., Hill, R. W., Kvien, C. K., Hamrita, T. K., & Rains, G. (2001). The Peanut Yield Monitoring System. Transactions of the ASAE, 44(4), 775–785.

Wakeley, P. C. (1954). Planting the southern pines. Washington, DC: USDA, Forest Service.

Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2022). ByteTrack: Multi-Object Tracking by Associating Every Detection Box. Proceedings of the European Conference on Computer Vision (ECCV).

Zukiwsky, N. M., Girard, T. E., & Zuidhof, M. J. (2020). Effect of an automated marking system on aggressive behavior of precision-fed broiler breeder chicks. Journal of Applied Poultry Research, 29(4), 786–797. https://doi.org/10.1016/J.JAPR.2020.06.005

Chapter 2

IN-FIELD PINE SEEDLING COUNTING USING END-TO-END DEEP LEARNING FOR INVENTORY MANAGEMENT

Published in Journal of the ASABE 66(2): 469-477 (doi: 10.13031/ja.15383).

Rafael Bidese-Puhl[1], Yin Bao[1], Nina Payne[2], Thomas Stokes[2], Ryan Nadel[2], Scott Enebak[2]

[1] Department of Biosystems Engineering, Auburn University, Auburn, Alabama, USA

[2] College of Forestry, Wildlife and Environment, Auburn University, Auburn, Alabama, USA

Highlights

- A deep learning-based machine vision system was developed for pine seedling counting.
- Automated seedling counting achieved less error than the manual sampling-based practice.
- Regression-based counting from optical flow inputs achieved the best performance.
- Machine counting produces seedling density maps for management practice improvement.

Abstract: *The southern U.S. produces over 1 billion pine seedlings for market sales per year, with prices varying from 50 to 435 dollars per thousand seedlings. An accurate inventory of seedlings provides nursery management with insights into how many seedlings can be sold and/or if there is any loss due to washout, mechanical damage, or pest/diseases that can still be mitigated. In this study, we developed a system to count pine seedlings at production sites and map the seedling density in the field. A system with three cameras was developed to collect video from different drills in the seedling bed. The videos were preprocessed to restrict the region of interest to the center portion of the image in each camera and separated each drill into individual videos. Two different modalities, i.e., video and optical flow, were evaluated as inputs to a convolutional neural network followed by a long*

short-term recurrent network to model the sequence of frames and regress to the seedling count for each plot. The mean absolute percentage error (MAPE) of our best performing model was 7.5%, which is an improvement over the baseline manual sampling-based approach with a MAPE of 11%. The results showed that the proposed approach was able to count seedlings in a crowded scene under complex field conditions with higher accuracy than the standard manual practice. Therefore, the proposed system and results demonstrated the potential to replace manual counting and even provide further information such as a seedling density map over the field for precision forest nursery management and seedling harvesting.

Keywords: CNN, LSTM, Nursery Inventory, Optical Flow, Pine Seedling, Regression.

2.1 Introduction

Pine plantations cover 16 million hectares (84%) of the total planted forests across the Southern United States (Miles, 2019). To meet the demand for reforestation, the Southeastern U.S. produces over 1 billion seedlings per year (Fargione et al., 2021). The state of Alabama alone is responsible for 91 million seedlings, ranking as the 4th largest producer in the U.S. (Haase et al., 2020). Recent efforts in improving seedling genetics have increased the value of individual seedlings, ranging from $50 to $435 per thousand seedlings depending on the pollination process used to guarantee the improved genetics (e.g., open pollination, $50-$75; mass control pollination, $120-$140; and somatic embryogenesis, up to $430) (Rousseau et al., 2012). With an increase in crop values in the field, it is paramount for forest tree nurseries to obtain accurate seedling inventories.

It is a common practice in forest nurseries to gather seedling inventories twice per year (in spring and fall) to determine the number of seedlings that are available and can potentially be sold. The spring seedling inventory is necessary for nursery managers to determine seedling bed densities and update seedling losses due to washout, mechanical damage, and pest/disease issues. This inventory is the first assessment of crop development and aids managers in determining areas of concern and where mitigation can still occur. The

fall inventory is more important than the inventory in spring, as it is needed to estimate the number of shippable seedlings and anticipate final sales.

The widely used seedling counting practice today was established more than half a century ago (Wakeley, 1954). It requires selecting multiple 0.3 m $\times$ 1.22 m (1- by 4-foot) plots for each seed lot used at the nursery. For each sample plot, the total number of seedlings is manually counted. A worker must bend over, sort through densely populated seedlings, and count the narrow stems that are growing within the counting frame placed on the bed. The inventory task is labor-intensive, time-consuming, error-prone, and ergonomically poor. Commercial nurseries have hundreds of kilometers of beds to cover with a limited staff. Therefore, automated seedling inventory technology can significantly reduce labor costs and improve working conditions for forest tree nursery production.

The first forest tree seedling counter was developed by the USDA Forest Services Missoula Technology and Development Center (Gasvoda and Herzberg, 1993). A side-viewing infrared optical sensor was used to detect seedling stems. The machine counting was within $\pm$10% of the manual counting. However, due to the side-viewing mechanism, large drill row spacing (i.e., 0.152 m) was required to carry the optical sensor. This system cannot operate on typical modern commercial seedling beds where there is no space to position a side viewing sensor. To the best of our knowledge, there is no other related work in the literature.

Recent advancements in computer vision and deep learning (DL) with the development of efficient algorithms like convolutional neural network (CNN) have enabled new opportunities for plant phenotyping (e.g., automated architectural traits extraction, and organ detection and counting) (Li et al., 2020). There are multiple ways in which the deep learning task can be formulated depending on the different objectives and data modalities (e.g., static images and videos) (Jiang and Li, 2020). For counting in static images, the main tasks can be formulated as regression, object detection, and semantic segmentation. Regression has been widely used for counting plants in high density, such as wheat (Lu and Cao, 2020),

rice (Liu et al., 2020), and maize (Lu et al., 2017). For counting in videos, the image sequence adds another level of complexity of associating the same objects across consecutive frames. This problem is typically solved by using tracking algorithms or by reconstruction of the scene in three-dimensional (3D) space (Sun et al., 2020; Tan et al., 2022).

Karpathy et al. (2014) demonstrated a video classification pipeline using pre-trained CNNs for feature extraction and different embedding fusing techniques and improved the state-of-the-art (SOTA) by 20% on the UCF-101 dataset (Soomro et al., 2012). Inspired by the brain processing of action identification, Simonyan and Zisserman (2014) explored the concept of combining optical flow and RGB videos for action recognition. They used a spatial CNN and a temporal CNN and later concatenated their outputs and fed them to a fully convolutional layer to perform classification with an improvement of 10% over Karpathy et al. (2014). Zhao et al. (2020) added a Long Short-Term Memory (LSTM) network (Hochreiter and Schmidhuber, 1997) to the model, using different approaches to fuse the outputs from the CNNs and the LSTM. The best performing model was a two-stream model using DenseNet (Huang et al., 2017) backbones pre-trained on the ImageNet dataset (Deng et al., 2009), which improved the SOTA accuracy for action recognition in the UCF-101 to 92.5%.

Similarly, other authors have proposed regression from videos for counting people in crowded scenes. Xiong et al. (2017) used convolutional LSTM networks (Shi et al., 2015) capturing both spatial and temporal cues. This work reported their mean absolute error (MAE) for different benchmark datasets. They achieved MAEs of 2.24 for the Mall dataset (Chen et al., 2012), 7.1 for scene 1 in the WorldExpo dataset (Zhang et al., 2016), and 1.3 for the UCSD dataset (Chan et al., 2008). Wu et al. (2019) obtained better results than Xiong et al. (2017) by employing a Temporally Aware Network building block with a novel lightweight convolutional network architecture. For the same datasets, they achieved MAEs of 2.03, 2.8, and 1.08, respectively.

The U.S. bareroot pine seedling production system employs a high planting density (e.g., 269 seedlings per square meter). In fall, when the seedling canopy covers the entire seedling bed, it is impossible for humans to visually separate individual seedlings from a top-view while standing on the ground in a non-contact manner, let alone for unmanned aerial vehicles with a high-resolution camera. We employed a push bar that bends the seedlings downward to expose the individual seedling stems. As the push bar moves forward, seedlings are released and return to their upright position. The movements of those seedlings are videotaped and used as information to perform the seedling count. A typical forest nursery can have 1.6 million seedlings per ha, making counting every seedling impossible.In this work, contrary to other counting systems that employ the detect-and-track approach, we present a machine vision system to count pine seedlings using end-to-end deep learning regression based on a CNN coupled with LSTM. The specific research objectives were to: (1) develop a deep learning-based machine vision system to count mechanically pushed pine seedlings, (2) simulate the accuracy of the standard manual sampling-based seedling counting, and (3) evaluate the machine counting performance against manual ground truth counting and the standard method.

2.2 Materials And Methods

2.2.1 Field Data Collection

Field data were collected at a commercial forest tree nursery in Georgia, USA, on 11 November 2021. The nursery name and location are not described here due to privacy reasons. The nursery consists of multiple units, and each unit is planted with single or multiple seed lots. A visual description of the field where the data were collected is presented in Figure 2.1. There are 9 beds in a unit, and a bed is approximately 183 m in length and 1.2 m in width. Loblolly pine seedlings are sown in 8 drills, with each drill row comprising two closely spaced rows in an alternating pattern to minimize seedling competition. Evenly spaced linear plots were selected in beds 2, 5, and 8 in Unit 1 and bed 2 in Unit 2 to increase

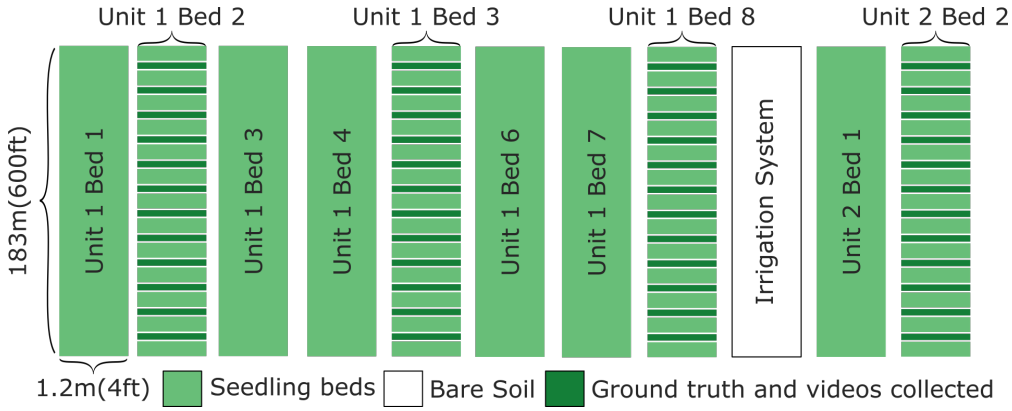


Figure 2.1: Description of field where data were collected. Seedling beds are 1.2m (4ft) wide and 183m (600ft) long. Each plot is 0.6 m in length and isolated by removing seedlings within 0.3 m before and after plot boundaries. Video and ground truth data were collected at dark green regions for evaluation.

the sampling area in contrast to sampling four consecutive beds. In each bed, 12 1.2 m $\times$ 0.64 m plots were separated by removing the adjacent seedlings outside of each plot (i.e., approximately 0.3 m per side) to enable clear distinction between the sample plots where ground truth was collected and the bed. Note that the current seedling counting practice employs a counting frame with a size of 0.3 m $\times$ 1.2 m, and our plot size was twice as large as a typical counting frame, resulting in an average stand count of 269 seedlings per plot. The stand count in each drill of each plot was manually obtained as the ground truth data. In total, the 48 plots resulted in approximately 13,000 seedlings.

2.2.2 Manual Counting Data Simulation

To evaluate the possible errors from the manual counting practice, the ground truth stand count data was used to simulate the errors that a nursery manager could make by using the current manual sampling-based counting method. Specifically, all possible combinations of selecting one plot from each bed were generated, and the average stand count of the four plots for each combination was used to estimate the seedling count for all plots. Depending on the combination of plots that are picked, the error in estimating plot seedling counts could be different. For comparisons, the combination of ground truth plots that give a result

close to the mean absolute percent error (MAPE) of all the possible estimations was used to represent the mean performance of the manual counting method. The same average value of the four plots was used to predict unit-level counts and was evaluated using MAPE. This simulation method was designed based on the fact that the current practice in the industry is typically counting seedlings in one counting frame per bed and using the measurement to estimate the total stand count for one or multiple beds depending on available labor.

2.2.3 Tractor-Based Video Collection Of Mechanically Exposed Pine Seedlings

The video acquisition system was a tractor implement (Figure 2.2a) that carried three wireless RGB cameras (Virb Ultra 30, Garmin, USA). The implement was attached to the three-point hitch of a John Deere 6110 utility tractor (John Deere, USA). The cameras were positioned at a nadir angle and configured to record videos at 30 frames per second with a resolution of 3840 $\times$ 2160 pixels. The cameras placed on the outer sides of the tractor implement were positioned to image the two outer drills of the bed, and the center camera was directly above the three middle drills (Figure 2.2b). Even though the cameras had a wide field of view that covered more than three drills, only the drills in the middle of the image were used mainly due to more occlusion occurring at drills further away from the center due to the slanted viewing angles.

The RGB cameras used offer an application program interface (API) to remotely control the cameras (e.g., start recording and stop recording). The API is based on hypertext transfer protocol (HTTP) requests, which require the user to be connected to the same network as the cameras; therefore, the cameras were configured to automatically connect to a wireless router placed in the tractor implement. The operator inside the tractor cabin used a laptop to run a custom-built graphical user interface (GUI). This GUI was developed to visualize the live feeds of the cameras, start recording, stop recording, and monitor video camera parameters (i.e., battery, storage, and operation status) (Figure 2.3). Both start and stop buttons on the interface broadcast calls to the APIs of all cameras, meaning that they

are operated remotely and simultaneously. To alleviate low light and shallow conditions under natural lighting conditions, a consumer-grade 1.3 m LED light bar (OffRoadTown, Guangzhou, China) was used to illuminate an imaging area of 0.3 m × 1.2 m on the seedling bed. A rolling push bar made of a PVC pipe with a 76 mm diameter was used to bend the seedlings down while the tractor traveled through the field at 0.65 m/s (i.e., 1.5 miles per hour). Note that the push bar does not damage the seedlings, because they are flexible enough to be bent by 90 degrees and recover their original upright positions as soon as they are released. Effectively, the push bar exposed individual seedling stems to the cameras. Additionally, the movement of each seedling provides another source of information about the density (i.e., more movements, more seedlings).

2.2.4 Video Pre-Processing

First, the RGB videos of the seedling beds were trimmed into short video clips of individual plots, with the beginning and end showing the empty gaps. Subsequently, because multiple seedling drills were visible in the video frame, each drill was cropped from the center of the frame with a 20% side overlap, generating a set of 8 video clips per plot. All videos were inspected to guarantee that the drill numbering was correctly assigned and that the cameras did not move sideways. Side movement was minimal due to narrow tractor tracks, straight beds, and tight adjustment in the three-point hitch holding the implement. This step reduced the input image size to the deep learning model and maintained the image resolution. The cropped video had a resolution of 852 × 720 pixels and a maximum length of about 3 seconds (i.e., 90 frames), resulting in a final dataset with 384 videos.

In addition to the RGB videos, the corresponding optical flow (OF) videos were used as input for the machine learning model. OF computes the distribution of apparent velocities of movement of brightness patterns in an image. For instance, Figure 2.4a shows an RGB frame where there is only one seedling moving upwards, and Figure 2.4b shows the respective OF for that frame. Since it is generated from the relative motion between the objects and

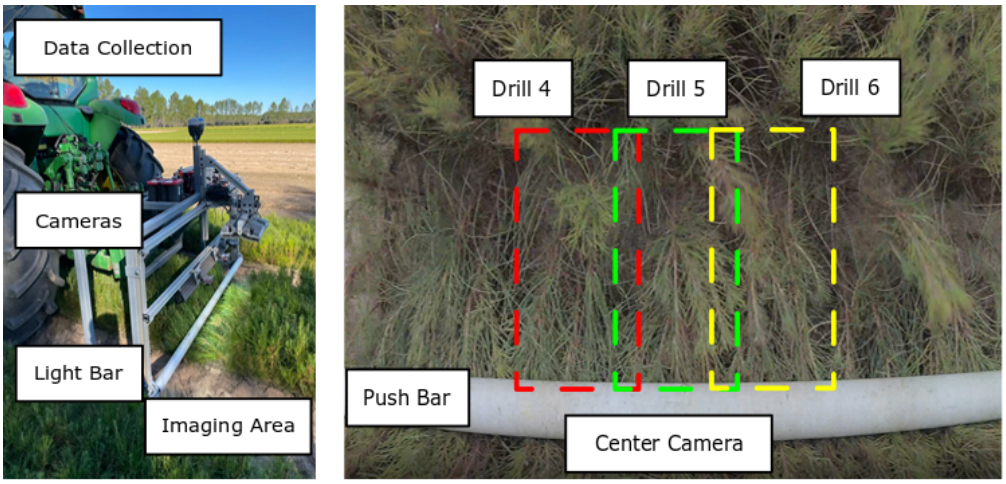


Figure 2.2: (a) Tractor-carried video acquisition system with LED light bar, three RGB cameras and push bar. (b) Sample video frame from center camera showing regions of interest in Drill 4, 5, and 6.

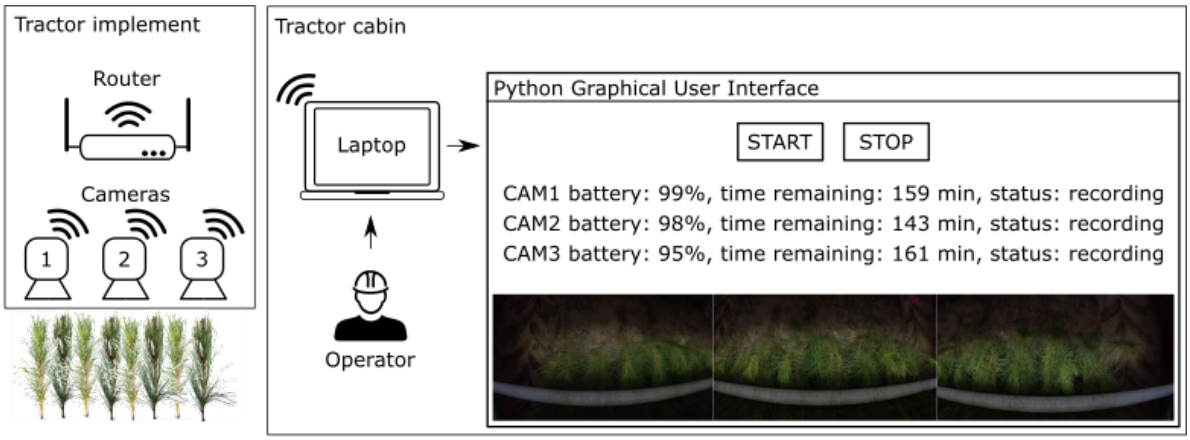


Figure 2.3: Data collection interface and controls diagram. Operator controls cameras using laptop connected to same wireless local area network as cameras.

the camera, it can give valuable information about the spatial arrangement of the objects and the rate of change of this arrangement (Horn and Schunck, 1981). The dense optical flow was computed using OpenCV 4.5.5 (Bradski, 2000) implementation of the Dual Total Variation L1 (Dual TV-L1) algorithm (Sánchez Pérez et al., 2013) and Python 3.8.12 (van Rossum and Drake, 2009).

2.2.5 Convolutional Neural Network With Long Short-Term Memory (CNN+LSTM)

The proposed deep learning model architecture consisted of a CNN followed by an LSTM and fully connected layers (Figure 2.5). Either an RGB video (Figure 2.5a) or the corresponding optical flow video (Figure 2.5b) can be used as the input, which was inspired by Zhao et al. (2020). This architecture was originally developed to perform human action recognition in videos as a classification task. In this study, we modified the architecture to perform regression. The neural network first uses a CNN to extract features from either an RGB or OF video frame and then feeds the feature embedding to an LSTM. The LSTM captures the temporal information in the sequence of high-level features extracted from the inputs, in this case, the movement of the bent seedlings springing back to their standing position. The output of the LSTM network is then processed by one fully connected layer (FC) to predict seedling count. The dimensions of the FC match the output of the LSTM. If the LSTM is one-directional, the FC input dimension is the size of the hidden layers divided by the number of layers. And, if it is a bi-directional LSTM, the outputs of both the forward and backward directions are concatenated, therefore doubling the size of the input for the FC.

Model training and evaluation were conducted using a dataset split based on a modified k-fold validation scheme, where each model was trained on three beds and evaluated in a fourth bed, for a total of four folds. This was done for all possible combinations for each model, and then the results for all folds were averaged and reported.

Parameter	Values	Training
Backbone	ResNet18, ResNet34, ResNet50, MobileNetV2, MobileNetV3Small, EfficientNetB0, EfficientNetB1	EfficientNetB0
LSTM Layers	1, 2, 3, 4, 5	3
LSTM Hidden Units	64, 128, 256, 512	512
Batch Size	1, 2, 4, 8	4
Learning Rate	0.1, 0.01, 0.001, 0.0001	0.01
Bidirectional	True, False	True
CNN Pre-trained	True, False	True
CNN Freeze	True, False	False

Table 2.1: Grid search hyperparameters and final training parameters. "Values" represent all hyperparameters that were evaluated, and "Training" represents the set of hyperparameters that was used for reporting results.

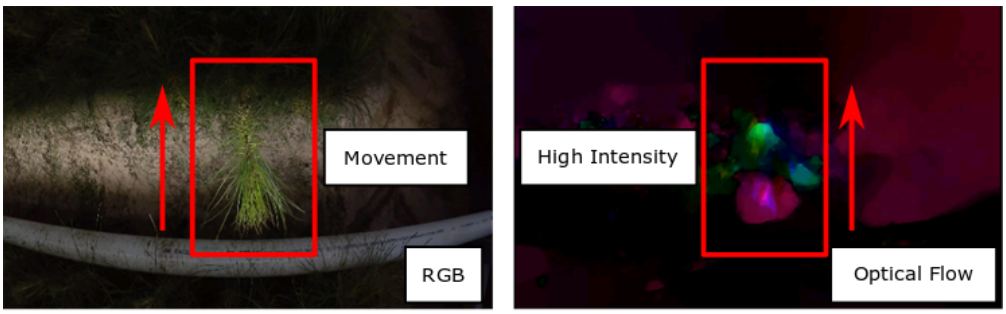


Figure 2.4: (a) RGB frame of pine seedling moving to upright position after released by push bar. (b) Correspondent optical flow frame with higher color intensity in fast-moving region of frame.

Different pre-trained and not pre-trained CNNs were evaluated as feature extractors with RGB and OF streams as inputs using drill-level stand count prediction MAPE, including ResNet18, ResNet34, ResNet50, MobileNetV2, MobileNetV3Small, EfficientNetB0, and EfficientNetB1. For the backbone selection, a one-directional LSTM with three layers and 256 hidden units was used. The remaining hyperparameters for this procedure were a batch size of two and a learning rate of 0.001. EfficientNetB0 was found to achieve the best performance for both input modalities. Subsequently, hyperparameter tuning was done using EfficientNetB0. A grid search was performed to determine the final parameters (Table 2.1).

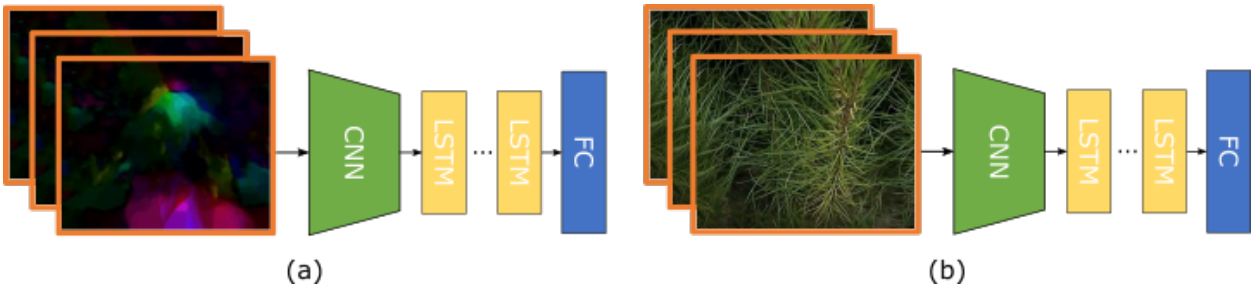


Figure 2.5: Convolutional Neural Network (CNN) with Long Short-Term Memory Network (LSTM) and Fully Connected (FC) regression architecture. (a) Using optical flow as input to model. (b) Using RGB frames as input to model.

For the LSTM, the hyperparameters included the number of layers (1, 2, 3, 4, 5), the number of hidden cells (64, 128, 256, 512), and the information flow direction (bidirectional or unidirectional). For the feature extractor, the hyperparameters were whether a pre-trained model was used or not and if the parameters of the network were fixed during training or not. For all hyperparameters, the fully connected layer matched the size of the output of the LSTM model, meaning the number of LSTM layers and hidden units as well as if it was bidirectional or not. Therefore, the final full network used 3072 input features and one output. For the whole network, the batch size (1, 2, 4, and 8) and the learning rate (0.1, 0.01, 0.001, and 0.0001) were evaluated.

The neural network was implemented using PyTorch 1.11.0 (Paszke et al., 2019) and CUDA 10.2 (NVIDIA et al., 2020) and trained for 30 epochs using the Adam optimizer (Kingma and Ba, 2015) with a scheduled learning rate decay of 0.1 every 10 epochs. The mean square error (MSE) was used as the loss function, and the best performing model was saved for evaluation. The model training was conducted on a workstation with an AMD Ryzen Threadripper 24-Core Processor with 64GB memory and an NVIDIA Titan RTX GPU with 24GB memory.

2.2.6 Model Performance Evaluation

Because the seedling bed videos were separated into individual seedling drill videos, the network was trained and tested on individual drill videos. During the evaluation, all drills

of a plot were evaluated, and the results were summed to get the estimated seedling count at the plot level. Given the two input modalities, three models were evaluated: RGB video as input to the CNN+LSTM network (i.e., RGB_CNN+LSTM), OF video as input to the CNN+LSTM network (i.e., OF_CNN+LSTM), and an average of RGB_CNN+LSTM and OF_CNN+LSTM. MAPE was used to evaluate the performance of the models. Evaluations were conducted for two different scales of measurement (plot level and unit level), using MAPE. Additionally, the machine-derived stand counts were compared to the average result that would be obtained by the nurseries if they had used the manual sampling-based counting method (Wakeley, 1954).

2.3 Results

2.3.1 Manual Counting Statistical Analysis

Using the ground truth data collected, all possible outcomes of the manual sampling of the seedling densities in the field were simulated to estimate a constant linear density of seedlings across the unit. The distribution of the errors for the different combinations to sample the field as well as the MAPE for such processes is presented in Figure 2.6. The unit error distribution is centered on 0%, whereas the plot prediction resulted in a skewed distribution with more extreme errors. The MAPE was 3.7% for unit-level prediction and 11% for plot-level prediction. However, the errors had a wide range, from -12% to 18% for the unit-level prediction and from -60% to 30% for the plot-level prediction. The simulation results show that there is clear spatial variability in seedling density within the beds. As a result, sampling location can have an impact on inventory estimation.

2.3.2 Machine Counting Performance

The model-predicted stand count results were evaluated using MAPE and the coefficient of determination (R^2) for plot and unit scales (Table 2.2). In addition, the results from the machine-based counting were compared to the manual counting simulation. The

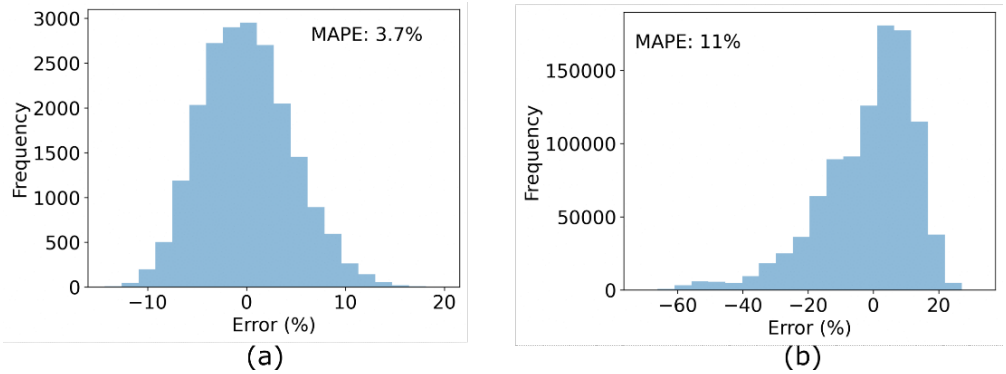


Figure 2.6: Histogram of simulated errors of manual sampling-based seedling inventory practice when used to estimate (a) unit-level and (b) plot-level inventory for a nursery in Georgia (USA).

sampling-based counting practice resulted in a plot-level MAPE of 11% and a unit-level MAPE of 3.7%. All three CNN+LSTM models had smaller MAPEs at plot and unit levels in comparison to the baseline counting practice. The average of the RGB_CNN+LSTM and OF_CNN+LSTM models achieved the highest R^2 of 0.42, with a plot-level MAPE of 8.3%. However, OF_CNN+LSTM achieved the lowest plot-level MAPE of 7.5%, a unit-level MAPE of 1.4%, and an R^2 of 0.41. Finally, RGB_CNN+LSTM resulted in a plot-level MAPE of 9.4%, a unit-level MAPE of 3.2%, and an R^2 of 0.33. The best deep learning-based model improved the current plot-level error from 11% to 7.5%, representing an improvement of 3.5%, and in a unit-level from 3.7% to 1.4%, or an improvement of 2.3%.

	Metric	**Sampling-based Counting**	**Deep Learning**		
			Optical Flow	**RGB**	**RGB + Optical Flow**
Plot	R^2	-	0.41	0.33	0.42
	MAPE (%)	11	7.5	9.4	8.3
Unit	MAPE (%)	3.7	1.4	3.2	2.3

Table 2.2: Counting results for different input streams. Counting column refers to the mean counting result when the method from Wakeley (1954) is used.

In addition to the previously mentioned data, the results can be visualized by evaluating the actual against predicted seedling counts per plot for the different models (Figure 2.7). OF_CNN+LSTM (Figure 2.7a) had the highest slope (0.36), and lower intercept

(188.6). While RGB_CNN+LSTM (Figure 2.7b) had the highest intercept (246.4) with the lowest slope (0.18). Also, being an average of the previous two, the (OF_CNN+LSTM) + (RGB_CNN+LSTM) (Figure 2.7) also had an intermediate intercept of 217.5 and a slope of 0.27.

2.3.3 Spatial Density Mapping

From the predicted counts and their respective spatial locations in the field, spatial mapping of the seedling count or seedling spatial density map can be obtained. Figure 2.8 shows the spatial maps of the stand count for ground truth (Figure 8a), the sampling-based counting method (Figure 2.8b), the OF_CNN+LSTM model (Figure 2.8d), the RGB_CNN+LSTM model (Figure 2.8e), and the average of the two DL models (Figure 2.8f), as well as their respective error maps (Figure 2.8c and 2.8g-i).

The ground truth map shows that the spatial variability in stand count for different beds in the field has significantly different seedling densities. A one-way ANOVA test of stand count between the four beds concluded that those differences were significant (p-value < 0.0001). The minimum and maximum values were 193 and 342 seedlings per plot, respectively. The sampling-based counting method achieved a constant count per plot across the field, and the errors varied from undercounting by 85 to overcounting by 63. The DL models were evaluated using 4-fold cross-validation by the bed. OF_CNN+LSTM achieved a maximum overcounting error of 43 seedlings and a maximum undercounting error of 67 seedlings. RGB_CNN+LSTM resulted in a maximum overcounting error of 48 seedlings and a maximum undercounting error of 92 seedlings. Finally, averaging the two DL models led to a maximum overcounting error of 46 seedlings and a maximum undercounting error of 79 seedlings. Again, averaging RGB_CNN+LSTM and OF_CNN+LSTM did not outperform OF_CNN+LSTM alone.

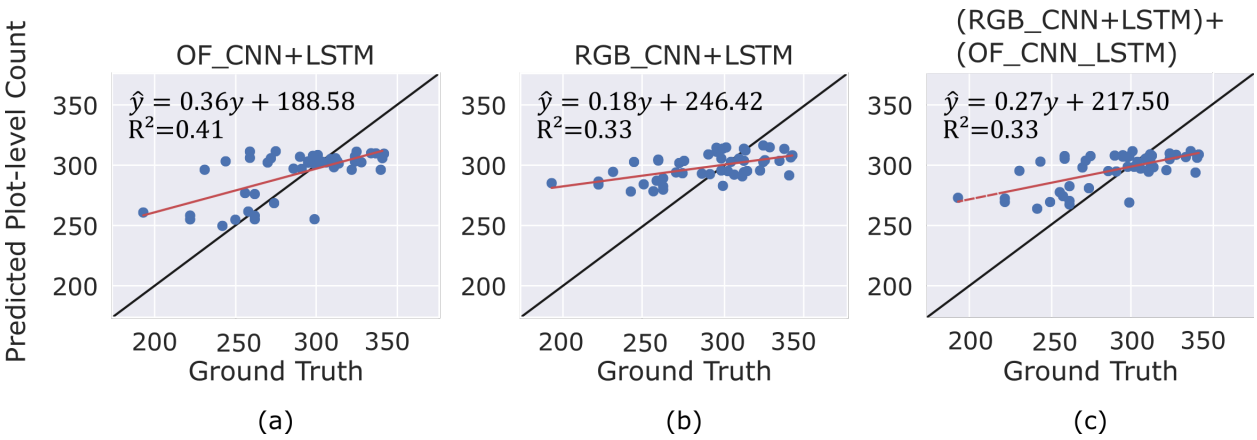


Figure 2.7: Scatter plots of actual vs predicted stand count at plot level. (a) Optical Flow (OF) as input for CNN followed by LSTM. (b) RGB as input for CNN followed by LSTM. (c) Combining results from (a) and (b) with average value.

2.4 Discussion

Taking inventory during pine seedling production is vital for nursery management, but it exposes the field workers to an unhealthy and demanding activity. Moreover, taking inventory more regularly is not possible due to the labor-intensive method used, which can take hundreds of person-hours to complete such a task using the current industry practice. Therefore, being able to automate this process with accurate counting can alleviate labor shortages and improve working conditions for field workers. In addition, automated seedling inventory is likely to increase profitability for nursery production by reducing labor costs.

By performing the manual sampling-based stand counting, a nursery manager would make an average error of 11% at the plot level and 3.7% at the unit level. However, the extremes of those events are significant. For instance, the unit-level error would range from -12% to 18%. The production could be overestimated by almost a third, which would provide inaccurate information for the nurseries to set their prices and estimate their earnings.

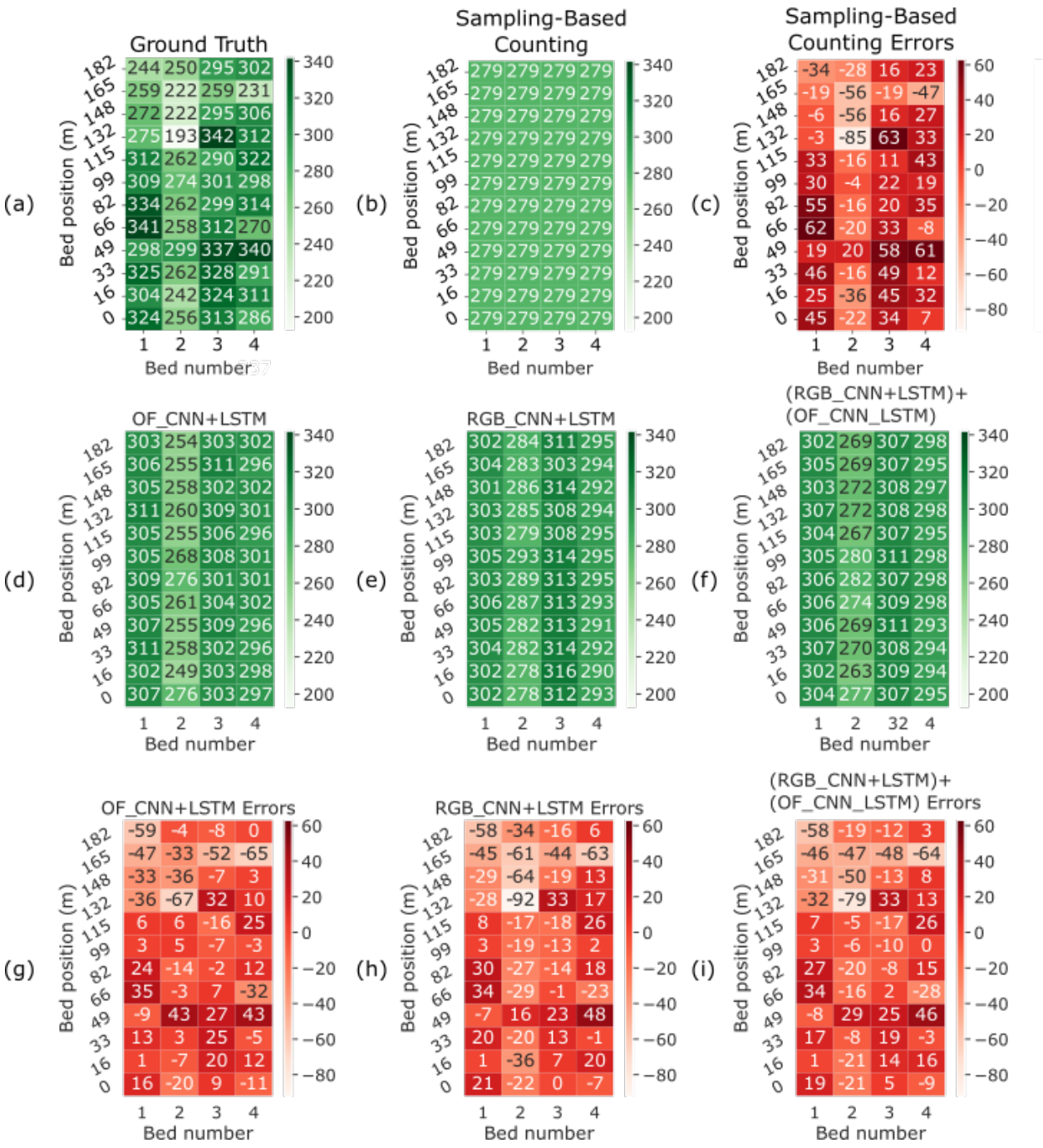


Figure 2.8: Comparisons of ground truth (GT), manual measurement, model-derived counts, and respective errors. (a) Ground truth, manually counted. (b) Average outcome derived from manual sampling-based counting. (c) Errors in estimating GT using manual sampling-based counting. (d) Predicted counts using optical flow as input for Convolutional Neural Network (CNN) followed by Long Short-Term Model (LSTM). (e) Predicted counts using RGB as input for CNN followed by LSTM. (f) Combining results from "d" and "e" with an average. (g) Estimation errors from model in "d". (h) Estimation errors from model in "e". (i) Estimation errors from model in "f".

The demonstrated deep learning models are trained in an end-to-end fashion. This approach requires less work (e.g., annotation and calibration) than its counterpart, component-based systems, which also demand a larger number of design choices. Moreover, the end-to-end model can potentially work for similar tasks by retraining the architecture with new data.

All the proposed methods performed better than the current counting practice in every metric evaluated. Figure 2.7 shows that the regression lines of the predicted plot-level count are more horizontal than the optimal (i.e., the diagonal line), which indicates that the models were not capable of extracting enough information to be able to predict counts that highly deviated from the average value accurately. It may be due to the dynamic and complex scene that has multiple sources of occlusion and fast-moving objects.

Further investigation into improving the feature extraction could improve the performance significantly since the pre-trained models were trained on out-of-domain datasets and the regression alone does not have strong labeling to help with learning. Pre-training the backbone on a similar task as seedling detection or segmentation could aid in the extraction of more relevant features. Moreover, self-supervised learning could make the pre-training model learn more relevant features. For example, as demonstrated by the Barlow Twins work (Zbontar et al., 2021), feature extractors can be trained to learn relevant features in a label-free manner. Experiments to further inspect and demonstrate which features are of interest to the neural networks to predict seedling count would help develop better architectures that can lead to performance improvements. Moreover, other stream fusion techniques such as early fusion, mid-fusion, and late fusion (Zhao et al., 2020) seem to have improved similar tasks in video classification and may have the potential to improve our model's accuracy as well.

The current counting practice does not give any insight into the spatial distribution of seedlings due to sparse sampling (e.g., one sample per bed). The proposed methods can generate a spatial distribution of seedling density with a smaller MAPE than the current

industry practice. It is important to notice that the stand count maps in Figure 2.8 have different mean bed stand counts, and the model was able to capture that. A result of the model's training with 4-fold cross-validation using each bed as a holdout for each fold demonstrates the ability of the model to generalize to different stand densities. Moreover, with the use of a GPS unit, the accurate field location of seedlings can be captured to generate spatially accurate maps of seedling density. Such a system can provide valuable insights for nursery managers to identify problematic areas for precision crop management and improve harvesting precision.

2.5 Conclusions

This study developed and evaluated a method for counting pine seedlings using videography of mechanically pushed seedlings and end-to-end deep learning. The proposed CNN- and LSTM-based regression network achieved MAPEs lower than the currently used manual sampling-based counting practice. Specifically, using optical flow as model input achieved the best performance with a MAPE of 7.5% at the plot level. But even the other models achieved superior performances when compared to the current practice MAPE of 11.07%. The model that used RGB video as input achieved a MAPE of 9.4%, and (RGB_CNN+LSTM) + (OF_CNN+LSTM) resulted in a MAPE of 8.3%. The experimental results show that the system could significantly increase the scale at which inventory on commercial pine seedling nurseries can be taken and therefore accurately harvest seedling orders. Moreover, the method shows the possibility of providing a spatial mapping of seedling density in the field that can be further used for decision-making in precision nursery management.

Acknowledgments

This project was supported by the Alabama Agricultural Experiment Station, the Hatch program of the U.S. Department of Agriculture National Institute of Food and Agriculture (USDA-NIFA), and the USDA-NIFA AFRI Foundational and Applied Science Program (Grant No. 2022-67021-37860).

The authors would like to thank Mary Beth Cassity, Nariman Niknejad, Kamand Bagherian, James Johnson, Bobby Bradford, Caroline Whiting, and Mery Margaret for supporting this project with ground data collection. The authors also thank James Johnson and Bobby Bradford for helping with system fabrication and equipment operation and the Southern Forest Nursery Management Cooperative nursery members at the College of Forestry, Wildlife and Environment at Auburn University.

REFERENCES

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

Chan, A. B., Liang, Z.-S. J., & Vasconcelos, N. (2008). Privacy preserving crowd monitoring: Counting people without people models or tracking. Proc. 2008 IEEE Conf. on Computer Vision and Pattern Recognition, (pp. 1-7). https://doi.org/10.1109/CVPR.2008.4587569

Chen, K., Loy, C. C., Gong, S., & Xiang, T. (2012). Feature mining for localised crowd counting. Proc. British Machine Vision Conference 2012, (pp. 21.1-21.11.). https://doi.org/10.5244/C.26.21

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. Proc. 2009 IEEE Conf. on Computer Vision and Pattern Recognition, (pp. 248-255). https://doi.org/10.1109/CVPR.2009.5206848

Fargione, J., Haase, D. L., Burney, O. T., Kildisheva, O. A., Edge, G., Cook-Patton, S. C.,... Guldin, R. W. (2021). Challenges to the reforestation pipeline in the United States. Front. For. Glob. Change, 4. https://doi.org/10.3389/ffgc.2021.629198

Gasvoda, D., & Herzberg, D. (1993). Seedling counter field test. Tree Planters' Notes, 44(1).

Haase, D. L., Pike, C., Enebak, S., Mackey, L., Ma, Z., & Silva, C. (2020). Forest nursery seedling production in the United States - Fiscal year 2019. Tree Planters' Notes, 63(2), 26-31. Retrieved from https://rngr.net/publications/tpn/63-2/forest-nursery-seedling-production-in-the-united-states2014fiscal-year-2019

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Comput., 9(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Horn, B. K., & Schunck, B. G. (1981). Determining optical flow. Artif. Intell., 17(1), 185-203. https://doi.org/10.1016/0004-3702(81)90024-2

Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. Proc. 2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), (pp. 2261-2269). https://doi.org/10.1109/CVPR.2017.243

Jiang, Y., & Li, C. (2020). Convolutional neural networks for image-based high-throughput plant phenotyping: A review. Plant Phenomics, 2020. https://doi.org/10.34133/2020/4152816

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014). Large-scale video classification with convolutional neural networks. Proc. 2014 IEEE Conf. on Computer Vision and Pattern Recognition, (pp. 1725-1732). https://doi.org/10.1109/CVPR.2014.223

Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. Proc. 3rd Int. Conf. on Learning Representations, ICLR 2015. https://doi.org/10.48550/arxiv.1412.6980

Li, Z., Guo, R., Li, M., Chen, Y., & Li, G. (2020). A review of computer vision technologies for plant phenotyping. Comput. Electron. Agric., 176, 105672. https://doi.org/10.1016/j.compag.2020.105672

Liu, L., Lu, H., Li, Y., & Cao, Z. (2020). High-throughput rice density estimation from transplantation to tillering stages using deep networks. Plant Phenomics, 2020. https://doi.org/10.34133/2020/1375957

Lu, H., & Cao, Z. (2020). TasselNetV2+: A fast implementation for high-throughput plant counting from high-resolution RGB imagery. Front. Plant Sci., 11. https://doi.org/10.3389/fpls.2020.541960

Lu, H., Cao, Z., Xiao, Y., Zhuang, B., & Shen, C. (2017). TasselNet: Counting maize tassels in the wild via local counts regression network. Plant Methods, 13, 79. https://doi.org/10.1186/s13007-017-0224-0

Miles, P. D. (2019). Forest Inventory EVALIDator web-application Version 1.6. 0.03. St. Paul, MN: USDA.

NVIDIA, Vingelmann, P., & Fitzek, F. H. (2020). CUDA, release: 10.2.89. Retrieved from https://developer.nvidia.com/cuda-toolkit

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G.,... Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, & R. Garnett (Eds.), Advances in Neural Information Processing Systems 32 (NeurIPS 2019) (Vol. 32). Curran Associates.

Rousseau, R. J., Roberts, S. D., & Herrin, B. L. (2012). Comparison of advanced genetic loblolly pine planting stock. In J. R. Butnor (Ed.), Proceedings of the 16th Biennial Southern Silvicultural Research Conference. Technical Report SRS-156. (pp. 269-272). Asheville, NC: USDA Forest Service, Southern Research Station.

Sánchez Pérez, J., Meinhardt-Llopis, E., & Facciolo, G. (2013). TV-L1 optical flow estimation. Image Processing On Line, 3, 137-150. https://doi.org/10.5201/ipol.2013.26

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., & Woo, W.-C. (2015). Convolutional LSTM Network: A machine learning approach for precipitation nowcasting. Proc. 28th Int. Conf. Neural Information Processing Systems, (pp. 802–810). https://doi.org/10.48550/arxiv.1506.04214

Simonyan, K., & Zisserman, A. (2014). Two-stream convolutional networks for action recognition in videos. Proc. 27th Int. Conf. Neural Information Processing Systems - Volume 1, (pp. 568-576). https://doi.org/10.48550/arxiv.1406.2199

Soomro, K., Zamir, A. R., & Shah, M. (2012). UCF101: A dataset of 101 human actions classes from videos in the wild. arXiv:1212.0402. https://doi.org/10.48550/arxiv.1212.0402

Sun, S., Li, C., Chee, P. W., Paterson, A. H., Jiang, Y., Xu, R.,... Shehzad, T. (2020). Three-dimensional photogrammetric mapping of cotton bolls in situ based on point cloud segmentation and clustering. ISPRS J. Photogramm. Remote Sens., 160, 195-207. https://doi.org/10.1016/j.isprsjprs.2019.12.011

Tan, C., Li, C., He, D., & Song, H. (2022). Towards real-time tracking and counting of seedlings with a one-stage detector and optical flow. Comput. Electron. Agric., 193, 106683. https://doi.org/10.1016/j.compag.2021.106683

van Rossum, G., & Drake, F. L. (2009). Python 3 Reference Manual. CreateSpace.

Wakeley, P. C. (1954). Planting the southern pines. Washington, DC: USDA, Forest Service.

Wu, X., Xu, B., Zheng, Y., Ye, H., Yang, J., & He, L. (2020). Fast video crowd counting with a Temporal Aware Network. Neurocomputing, 403, 13-20. https://doi.org/10.1016/j.neucom.2020.04.071

Xiong, F., Shi, X., & Yeung, D. Y. (2017). Spatiotemporal modeling for crowd counting in videos. Proc. 2017 IEEE Int. Conf. Computer Vision (ICCV), (pp. 5161-5169). https://doi.org/10.1109/ICCV.2017.551

Zbontar, J., Jing, L., Misra, I., LeCun, Y., & Deny, S. (2021). Barlow twins: Self-supervised learning via redundancy reduction. In M. Marina, & Z. Tong (Ed.), Proc. 38th Int. Conf. on Machine Learning, 139, pp. 12310-12320. Retrieved from https://proceedings.mlr.press/v139/zbontar21a.html

Zhang, C., Kang, K., Li, H., Wang, X., Xie, R., & Yang, X. (2016). Data-driven crowd understanding: A baseline for a large-scale crowd dataset. IEEE Trans. Multimedia, 18(6), 1048-1061. https://doi.org/10.1109/TMM.2016.2542585

Zhao, Y., Man, K. L., Smith, J., Siddique, K., & Guan, S.-U. (2020). Improved two-stream model for human action recognition. EURASIP J. Image Video Process., 2020(1), 24. https://doi.org/10.1186/s13640-020-00501-x

Chapter 3

AN MMWAVE RADAR-BASED MASS FLOW SENSOR USING MACHINE LEARNING TOWARDS A PEANUT YIELD MONITOR

Manuscript under review at *Computers and Electronics in Agriculture*

Rafael Bidese-Puhl[1], Christopher Butts[2], Matt Rewis[3], Joseph McIntyre[2], Jason Morris[3], Bennie Branch[3], Yin Bao[1]

[1] Department of Biosystems Engineering, Auburn University, Auburn, AL, USA

[2] National Peanut Research Laboratory, USDA-ARS, Dawson, GA, USA

[3] Kelley Manufacturing Co., Tifton, GA, USA

Highlights

- A novel millimeter-wave FMCW radar-based mass flow sensor for monitoring peanut mass flow rate during harvest was developed.
- Machine learning models accurately predict mass flow rate from radar data in both research and commercial settings.
- The sensor can provide real-time feedback for operator which can be used to reduce peanut shelling.
- The developed pipeline can be extended to mass flow rate measurement of other harsh environments in agriculture.

Abstract: *A millimeter-wave FMCW radar-based mass flow sensor was developed to monitor peanut mass flow rate during harvest. The radar sensor components can be placed outside the abrasive flow in the pneumatic conveyor of a peanut combine. Two systems to simulate the mass flow conditions of the field were used: one for research scale mass flow rate using a retrofitted 2-row combine blower and one for commercial scale mass flow rate using a modified 6-row combine. The ground truth for mass flow rate was obtained and*

the radar data used to get range-velocity image time-series. The datasets were generated using a sliding window interval, samples were generated to train machine learning models to predict mass flow rate from radar data. After evaluation of 5-fold cross validation, the best performing models achieved an RMSE of 0.14 kg/s (19 lb/min) and a sMAPE of 15 % while having an R^2 of 0.85 for the research-scale combine and an RMSE of 0.52 kg/s (69 lb/min) and a sMAPE of 10 % while having an R^2 of 0.71 for the commercial-scale combine. Moreover, the sensor can provide the combine operator with information about the velocity of the peanuts to adjust the air pressure of the pneumatic conveyor to reduce undesired peanut shelling. The mass flow measurement results are promising. Further field investigation is necessary to evaluate the effects of noise caused by combine movement, foreign materials, peanut varieties, moisture content, and soil type.

Keywords: Machine Learning, Mass Flow Sensor, Peanut, Radar, Yield Monitor.

3.1 Introduction

A yield monitor mounted on a combine, together with a global positioning system (GPS) allows the generation of yield maps that are used for identification of crop performance issues in the field that can be further investigated and possibly mitigated through management practices. Yield monitors for major crops such as corn, soybean, and cotton have already been widely commercialized, but many crops still do not have a reliable system to monitor harvest. Economic relevance is a major factor for why certain crops to have more developed technologies. Field conditions during harvest make the development of sensors for certain crop production systems even more challenging. During the harvest of peanuts, it is common for foreign materials (e.g. soil, vines, rocks, etc.) to run through the pneumatic conveyor in the combine. This makes the environment inside the air duct of the pneumatic conveyor harmful for components that are fragile. For instance, optical systems that have lenses or sensors inside the duct must employ strategies to reduce wearing of the parts due to the abrasive nature of the environment, but they are not enough to withstand this harsh

environment for long periods of time. Other load cells-based systems have been tested in the field but provide low spatial resolution, making management practices less accurate.

Some of the peanut yield monitors developed employed load cells mounted to the combine basket. Vellidis et al. (2001) developed a system for production scale with errors in the range of 5% for loads from 1,567 to 5,420 kg, but the yield map generated is noisy due to the movement of the combine as it traverses across the field. Kirk et al. (2012) developed a system capable of measuring cumulative weights of small research plots. Due to the low spatial resolution and high accuracy (Porter et al., 2020) suggests that load cell-based systems can be used as calibration checking for other systems. Optical-based systems were also developed by several works (Porter et al., 2017; Porter et al., 2020; Rains et al., 2005; Thomasson et al., 2006). In general, they require field-specific calibration and devise multiple techniques to extend working time before the internal optical components get damaged by the harsh environment of the pneumatic conveyor in the combine. Rains et al. (2005) after correcting measurements for moisture content and adding specially design deflector in the pneumatic conveyor to mitigate effects of foreign materials reported a mean absolute error in load weight of 3.2% to 9.4% for different fields using loads varying from 600 kg to 3,000 kg for evaluation. Thomasson et al. (2006) developed and tested an optical system adapted from cotton yield monitor using near-infrared light emitting diodes (Figure 3.1a). The optical system achieved R^2 up to 0.96 for field data harvesting loads of 500 to 3,000 kg, but problems with dirt buildup in days with high moisture content were reported. Moreover, this study performed tests for small batches (i.e., 25 kg) using different durations to run it through the sensor to evaluate linearity of the sensor and achieved R^2 of 0.84 when using the dumping time as reference for signal processing and an increased R^2 of 0.97 when the sensor readings were used to determine the dumping time. Porter et al. (2017) performed extensive field evaluation and calibration of the optical system, which demonstrated issues with damages in the lenses that deteriorated performance. After installing a deflector for small pebbles, the damage was reduced but still present during harvest such that compensation

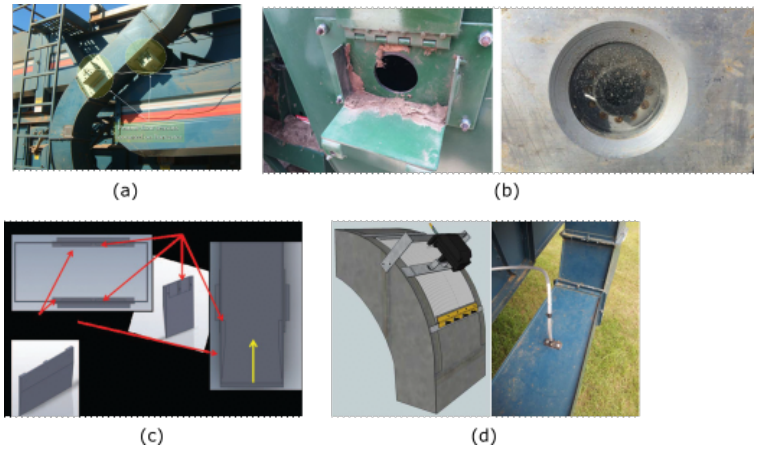


Figure 3.1: Existing approaches for peanut yield monitoring. (a) Optical system adapted from cotton yield monitor (Thomasson et al., 2005). (b) Extensive field evaluation of the optical system demonstrated issues with damages in the lenses that deteriorated performance (Porter et al., 2017). (c) Addition of a deflecting plate into the system to reduce damaging on the optical system (Porter et al., 2020). (d) Impact plate-based yield monitor using a compensation for air speed in the pneumatic conveyor (Free et al., 2014).

for certain data points were necessary up to 25% to generate yield maps (Figure 3.1b). The study presents results for different fields and demonstrated that there is a need to perform field-specific calibrations, otherwise the R^2 can be reduced to 0.80. Though, their individual field calibration had small number of samples to be able to generate a good prediction model. Porter et al. (2020) added a deflecting plate into the system using a retrofitted AgLeader cotton yield monitor sensor pair that achieved 11% error with a model calibrated for the same peanut variety (Spanish) and up to 25% for different peanut varieties (i.e., calibrated with Spanish and evaluated in Runner) (Figure 3.1c). Their system would report to the operator when the optical sensors had to be cleaned to maintain performance. Free et al. (2014) adapted an impact plate-based yield monitor for grains and compensated for the different air speeds in the air duct due to different field loads by using a pressure sensor in the pneumatic conveyor and achieved an R^2 of 0.83 and an mean absolute error of 6.6% with a standard deviation of 5.3% with 7 loads of 2,300 to 3,175 kg and one load of 1,600 kg (Figure 3.1d).

Downey et al. (2010) used different methods for specialty crops yield measurement (e.g., mass flow rate measurement, batch weighting, and cumulative weighing) are presented focusing mostly on different configurations of load cell systems. Though, more recently with the development and availability of modern technologies and perception sensors, different modalities of radio detection and ranging (radar) started to be incorporated into precision agriculture systems. Longchamps et al. (2022) reviews several modern approaches for yield estimation, including radar. Ultra-wideband (UWB) radar systems and ground penetrating radar have been used as a non-contact proximal sensor to estimate yield of a variety of crops that grow under the ground (potato, manioc, sugar beet). Konstantinovic et al. (2008) used UWB radar to detect the location of sugar beets. It was reported that there was an 80% correlation between the biomass of beets and the backscattered energy and above 90% accuracy in visual detectability of beetroot when using radargrams. It was also suggested that the method could be further improved using machine vision techniques to process the radar signal. Radar sensors have also been used in the agriculture for mapping plant vigor in vineyards. Henry et al. (2019) demonstrated a radar-based system that scanned the canopy to estimate the biomass of grapes. Although exploratory, the proposed approach detected most grapes in the scene of interest, even if the grapes were partially or totally occluded, hence overcoming the limitation of current optical sensors that require the visibility of all or parts of the objects of interest. Moreover, because of the small wavelength of the radar and the size of the dust particles in the environment, the radar signal would be minimally be affected by the presence of dust. Finally, the radar sensor can be packaged in a dustproof enclosure that can withstand the harsh harvesting environment.

Currently, there is no commercially available peanut yield monitor from peanut combine manufacturers and agricultural technology companies. To comply with such environment, a non-contact measurement of the yield is of need. Given the success of radar sensors in object detection (Henry et al., 2019; Konstantinovic et al., 2008), we propose to use a frequency modulated continuous wave (FMCW) millimeter-wave (mmWave) radar to measure mass

flow rate of pneumatically conveyed unshelled peanuts. A data-driven approach was investigated to model the relationship between the radar image signature to a given mass flow rate. To handle the high-dimensional radar range-velocity image, machine learning models are employed to perform the modeling task. The specific research objectives were to: (1) collect mmWave radar data from pneumatically conveyed peanuts at different mass flow rates, and (2) evaluate the performance of multiple machine learning models for prediction of peanut mass flow rate from the radar data.

3.2 Materials And Methods

Two systems were developed to evaluate commercial mmWave radar sensors for measuring mass flow rate of pneumatically conveyed peanuts. First, a system using a 2-row combine configuration to evaluate the sensor under low mass flow rate conditions to emulate research scale trials. Then, the second system was based on a modified commercial 6-row combine to evaluate the system under conditions closer to the on-farm application. The two systems employed different sensors and strategies for calibration, but shared the same modeling pipeline.

3.2.1 Data Collection Setup

2-row Combine

A pneumatic conveyor of a 2-row peanut combine was retrofitted to simulate varying peanut flows in an air duct during harvest at the USDA-ARS National Peanut Research Laboratory in Dawson, GA, USA. The pneumatic conveyor consisted of a blower fan (A4A, The Lima Electric Motor Co., Inc, USA), a variable frequency driver (VFD) (E-TRAC WFCHT Series, TB Wood, USA) for fan speed control, and an air duct. An overhead bin was placed on top of the frame that supported the pneumatic conveyor. A gate at the bottom of the bin was manually controlled to vary the amount of peanuts falling into the air duct. Different peanut mass flow rates were generated by changing the blower fan speed and the

gate opening. Peanuts coming out of the pneumatic conveyor entered a second bin on the floor. After the overhead bin was empty, the two bins were swapped using a forklift. The same batch of farmers' stock peanuts were reused during the experiment. A section of the air duct was replaced by an acrylic duct with the same internal dimensions (203.2 mm $\times$ 152.4 mm) as the air duct for the radar sensor to collect signals from the peanut flow, since the radio frequency signal can go through plastic but not metallic materials. An industrial floor scale (BT-FC-6060-10-CL, B-TEK Scales, USA) was placed under the system to measure the mass of peanuts leaving the system. The mass flow rate was calculated as the absolute rate of change of the scale measurements. A laptop (S410 G3, Getac, Taiwan) was connected to an mmWave radar (AWR2243BOOST, Texas Instruments, USA) via a data acquisition card (DCA1000, Texas Instruments, USA) to control the boards and record the raw readings from the radar. The same laptop was connected to the industrial scale via RS-232 to record the scale readings. Figure 3.2 shows both a diagram and a picture of the described system.

An industrial scale (BT-FC-6060-10-CL, B-TEK Scales, USA) was used to measure the ground truth mass flow rate. This scale provides an accuracy of 0.02% and works at a range from 1 to 3,535 kg. In addition to a display, the scale was configured to output the mass readings through a serial port using fixed length comma separated string at 10 samples per second. Those values were read and stored in a csv file for post-processing.

6-row Combine

A commercial 6-row peanut combine (Model 7406, Kelley Manufacturing Company, Tifton, GA, USA) was retrofitted to recirculate a peanut load between its tank and pneumatic conveyor. Specially, the tank was cut open at the bottom side and an electronically controlled conveyor belt was mounted under the tank to feed peanuts from the tank to the inlet of the pneumatic conveyor. The blower fan conducts the peanuts through the pneumatic conveyor back to the overhead tank. To acquire the radar data, a vertical section of the air duct was replaced by an identical part manufactured with a 5 mm acrylic sheet. The conveyor that

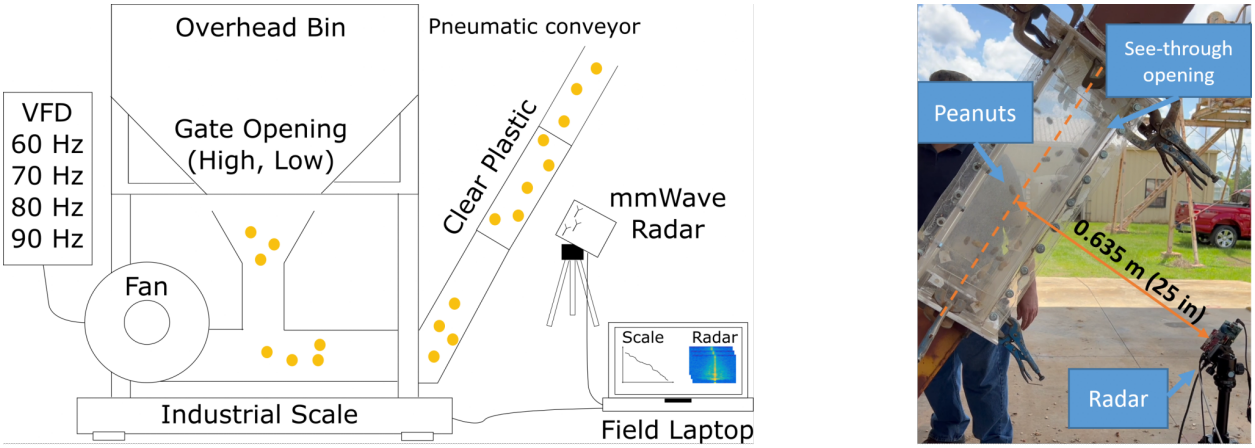


Figure 3.2: Data collection setup. A variable frequency drive (VFD) controls the fan speed; A gate controls the flow of peanuts into the air duct; an industrial scale is used to measure the peanut mass that is being discharged; the peanut flow through a plastic section in the mental air duct for the radar sensor to capture the peanut flow. A rugged laptop is used to control the radar and acquire data. (a) diagram (b) picture.

feeds the peanuts to the inlet of the duct was calibrated beforehand to produce a constant predictable mass flow rate. The calibration procedure was performed by placing a plastic container under the output of the conveyor belt. A laser sensor was installed in the container to detect when the peanuts start flowing to the container and when the peanuts stop flowing, allowing for accurate measurement of the flow duration. Before the container was full, it was removed from the conveyor belt output and weighted. To perform this calibration, 11 points were measured and then a linear regression was used to calibrate the relationship between conveyor speed and mass flow rate.

A standalone radar acquisition system was developed for this setup to serve as a yield monitor prototype (Figure 3.3). An compact radar development kit (IWR6843AOPEVM, Texas Instruments, USA) controlled by Raspberry Pi 4 Model B (Raspberry Pi Foundation, England) was employed for data collection (Figure 3.3b). The Raspberry Pi controls and reads the data from the radar through a Universal Serial Bus (USB) interface and records the data. The Raspberry Pi was configured to act as a Wi-Fi access point and serves a web interface (Figure 3.3a) that allows the operator to start, monitor, and stop acquisition of the

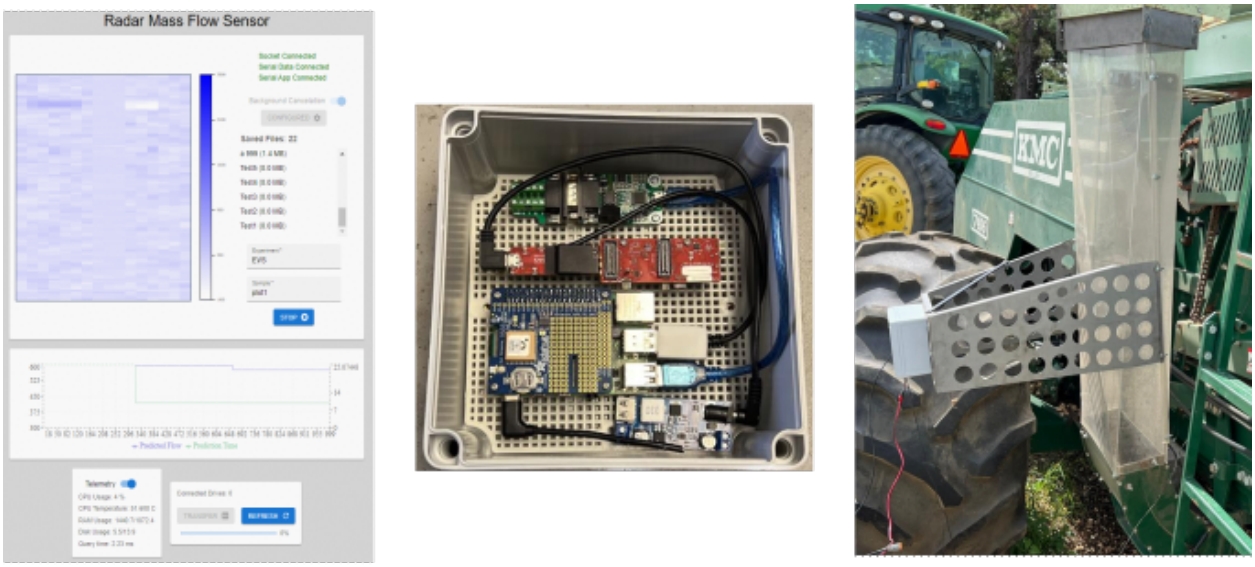


Figure 3.3: Prototype developed for 6-row combine experiments. (a) web interface. (b) radar system. (c) system mounted to the 6-row combine.

radar data, and transfer the collected data to a flash drive for further processing. Figure 3.3c shows the prototype mounted to the 6-row combine for data collection.

3.2.2 mmWave FMCW Radar

An mmWave radar sensor transmits frequency-modulated continuous waves (i.e., chirps) to the target environment using transmitting (TX) antennas, and the reflected signals from objects are captured by receiving (RX) antennas (Yadav et al., 2022). The specific parameters of those chirps translate into practical dimensions, ranges, and velocities that the sensor can measure. The chirp parameters shown in Table 3.1 were obtained by using the mmWave Sensing Estimator tool provided by Texas Instruments, which resulted in the scene parameters presented in Table 3.2. The chirp parameters were configured to maximize radial velocity without losing range resolution while maintaining a maximum radial velocity that is compatible with the expected scene. For both systems, the radar was placed pointing orthogonally to the duct at 0.635 m (24 in) from the duct.

Scene parameters	2-row combine	6-row combine
Maximum range (m)	4.96	12
Minimum range (m)	0.5	0.5
Range resolution (mm)	38.8	48.9
Maximum radial velocity (m/s)	8.6	8.3
Velocity resolution (m/s)	0.27	0.27

Table 3.1: Radar scene parameters for the two radar systems.

Parameter	2-row combine (AWR2243)	6-row combine (IWR6843)
Frequency range (GHz)	76 – 81	60 – 64
Bandwidth (MHz)	4537.65	4000
Frequency slope (MHz/µs)	90.33	96.37
Idle time (µs)	7	7
ADC samples	124	97
Sample rate (ksps)	3000	3000
Ramp end time (µs)	50.23	41.23
Chirps per frame	62	62
Frame period (ms)	10 (100 Hz)	200 (5 Hz)

Table 3.2: Radar chirp parameters for the two radar systems.

Radar Data Processing

The time domain signals collected by the radar were digitized and saved for off-line processing. The post-processing included performing sequentially a one-dimensional (1D) Fast Fourier Transform (i.e., range-FFT) to generate range-time images then a two-dimensional (2D) FFT (i.e., velocity-FFT) to generate range-velocity images. In a range-velocity image, the x-axis represents the radial velocity of particles, where a negative velocity means that an object is moving towards the radar and a positive velocity is an object moving away from the radar. The y-axis represents the distance that an object is from the radar. Moreover, the pixel values represent the strength of the signal that the radar receives from an object at a given distance moving at a given radial velocity. The x-axis spans over the possible velocities that the radar configuration can detect with steps of the size of the radar velocity resolution and similarly the y-axis for the range values and range resolution.

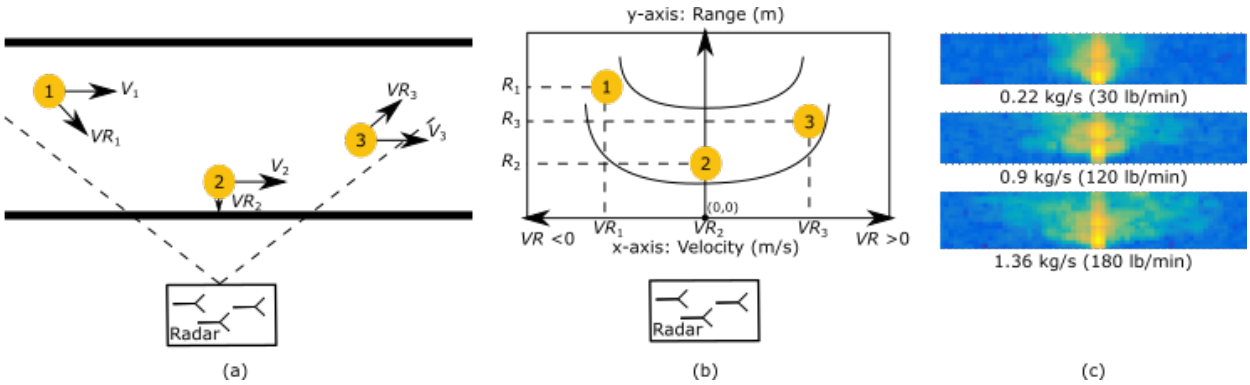


Figure 3.4: mmWave Radar working principle. (a) Scene inside the pneumatic conveyor. (b) Equivalent expected range-velocity image (c) examples of radar data collected for three different mass flow rates.

In this specific application, the radar data collection was arranged for sensing peanuts flying through the acrylic duct section of the pneumatic conveyor as shown in Figure 3.4. In Figure 3.4a, the duct with the radar positioned orthogonally to it can be seen. A peanut pod moving from left to right with a linear velocity V1 will be seen by the radar as an object with a radial velocity VR1 at a distance R1 relative to the radar as demonstrated in Figure 3.3b. As this object moves from the left side to the right side of the radar, it will also appear in a different position in the range-velocity map. For instance, if it follows a linear trajectory, it will follow a u-shaped trajectory in the range-velocity map. The intensity of the pixels on the range-velocity map is relative to the radio-frequency reflection from the object. When multiple objects are moving at the same time, the intensities of the pixels in the range-velocity image overlap each other and thus separating individual objects becomes infeasible. For reference, Figure 3.4c shows different mass flow rates and their respective range-velocity images collected on the 2-row combine system that will be described.

3.2.3 Data Collection

2-row Combine

A dataset was built by collecting radar and scale measurements at different levels of fan speed and peanut input gate opening. The blower fan was controlled with a VFD and

different values of frequency were used. Based on preliminary tests, 60, 70, 80, and 90 Hz output from the VFD were chosen as representatives of the range of air velocities for a 2-row combine harvesting peanuts in the field. To control the input of peanuts into the duct, the gate of the overhead peanut bin was used to control the peanut flow into two discrete relative amounts, low and high. The low peanut intake was set as to the smallest gate opening that the peanut flow would be constant, and the high peanut intake was set as the highest peanut flow that the fan speed could sustain without the peanuts accumulating in the air duct and breaking the flow. With the combination of fan speed and peanut intake, a wide range of uniformly distributed mass flow rates in the duct were generated. Each repetition of data acquisition lasted 2 minutes and is called a trial. Moreover, for each of the trial configuration, three separate acquisitions were done to ensure that the variations of the gate opening and the weight of the peanut column were not influential to the data. Therefore, given four fan speeds, two gate openings, and three repetitions, a total of 24 trials were conducted.

6-row Combine

A second dataset was built by collecting radar data at different conveyor belt speeds to adjust the mass flow rate of the 6-row combine system. Using the calibrated model for conveyor belt speed and peanut mass flow rate, different mass flow rate levels were chosen to collect the data based on the expected commercial field operation, ranging from 2.1 to 6.1 kg/s (279 to 811 lb/min) at an approximate increment of 0.11 kg/s (15 lb/min). The data collection process followed by starting the peanut mass flow. As soon as the flow became stable, the radar data acquisition was performed for 10 seconds using the web interface.

3.2.4 Data Pre-Processing

Data pre-processing for the two systems (2-row and 6-row) were different due to the different specifications of the two data acquisition systems. The radar sensor used in the 2-row system does not perform any onboard computing of the radio signals, while the one

used in the 6-row system computes the range-velocity image onboard. The ground truth for the 2-row system came from the industrial scale data, while the ground truth for the 6-row system comes from the calibrated conveyor belt feed system.

2-row Combine

The raw data from the radar system were pre-processed using the mmWave Studio (Texas Instruments, USA) to parse the radio-frequency signals collected during acquisition into a binary file. Subsequently, a MATLAB script available at mmWave Studio (Texas Instruments, USA) was used to convert the binary file into a standardized radar data format, a radar cube. Furthermore, the radar cube was processed using the OpenRadar library (Pan et al., 2019) to generate a video where the radar frames were converted into range-velocity images. A region of interest (ROI) was determined at the ranges from 0.4 m to 1 m given the minimum working distance of the radar and the location of the air duct relative to the radar. The scale and radar data were not synchronized during acquisition, Figure 3.5 shows the alignment pipeline employed in this work. After each trial started, the scale data acquisition was started before the radar acquisition. Once the 2-minute radar acquisition was completed, the scale acquisition was stopped. So, a lag needed to be added to the radar data so that it matches the scale data. To identify that lag, the radar video was down-sampled to match the sampling rate of the scale, the mean pixel value in the ROI of each frame was calculated to generate a 1D time-series that represented the range-velocity image. In this case, the mean pixel value acts as a hand-engineered feature to enable the data-driven alignment. Moreover, the derivative of the weight was calculated and since this operation can result into noisy signal, the derivative was further smoothened with a 1 second moving average. Finally, cross correlation was used to find the best alignment between the two data time-series. The position at which the absolute value of the correlation was highest was assigned as the best alignment between the mass flow rate and the range-velocity image.

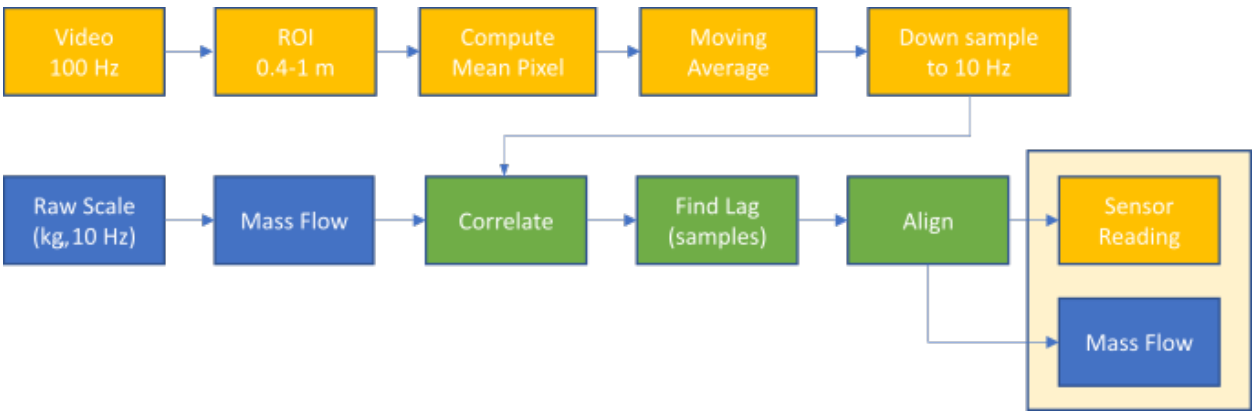


Figure 3.5: Alignment pipeline for the scale and radar data of the 2-row system.

There were two main sources of noise during the data acquisition. First, for the low flow gate opening configuration, sometimes the peanuts would stop flowing requiring the peanut bin to be shaken to force the peanuts to flow. Second, when attempting to get the maximum peanut flow from the blower fan the lift capacity would be exceeded and peanuts would accumulate in the duct. The trials that had excessive noise were removed from the dataset. Additionally, there were inconsistencies in the beginning and at the end of the trials, and the each trial dataset were trimmed to reduce noisy ground truth data.

3.2.5 Dataset

Two separate datasets were generated from the two systems using the same process after the mass flow rate and range-velocity data were obtained. To formulate a dataset for a supervised machine learning task, the mass flow rate data and range-velocity video were partitioned into pairs of samples using a 3-second sliding window with a 0.5-second overlap. Within each window, the average range-velocity image (X) and the average mass flow rate (y) were calculated. This dataset was further separated using a 5-fold cross validation scheme where 20% of the data was held out for model evaluation and the model was trained using the remaining 80% of the data. For the 2-row combine dataset, this resulted in 2416 training samples and 471 evaluation samples in each fold and for the 6-row combine dataset there were 228 training samples and 57 evaluation samples per fold. This dataset partition scheme

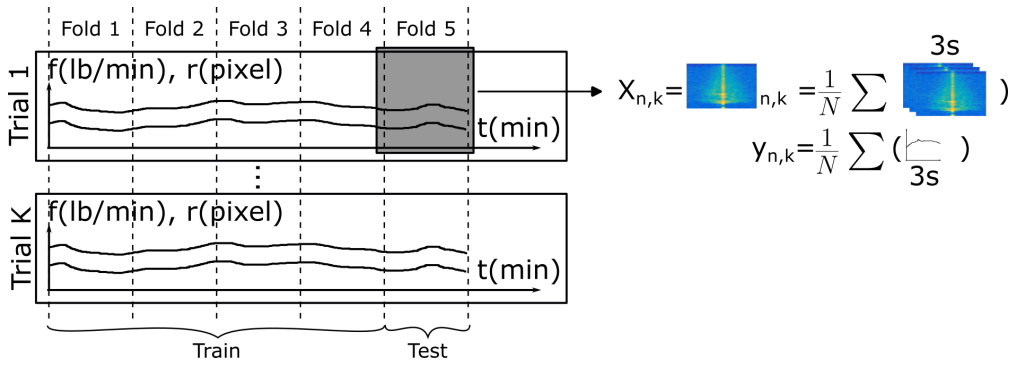


Figure 3.6: Dataset preparation. K is the number of trials, N is the number of samples, X is the input to the machine learning models, and y is the scale-derived mass flow rate.

ensured that the folds had a constant number of samples and evaluation across them was balanced. The dataset preparation is illustrated in Figure 3.6. Finally, the model evaluation results were averaged across all 5 folds.

3.2.6 Model And Evaluation

Different machine learning regression models were trained and evaluated. The models evaluated were linear regression, k-neighbors regressor, random forest regressor, support vector regressor with linear and radial basis function kernels, and a multi-layer perceptron regressor. As shown in Figure 3.7, the region of interest (ROI) on the range-velocity image is extracted based on the physical dimensions of the system, the ROI pixel values are flattened into a 1D vector, the models are trained to predict instantaneous mass flow rate and the results are evaluated. Background cancellation was also evaluated if it could improve the performance of the models. Background cancellation was implemented such that the average range-velocity image for zero mass flow rate was subtracted from the average image of each sliding window.

The models were evaluated by how well they can predict instantaneous mass flow rate using root mean squared error (RMSE), coefficient of determination (R^2) and symmetric mean average percent error (sMAPE).

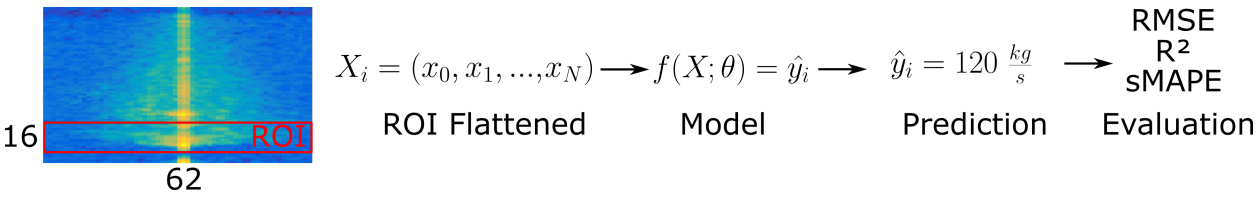


Figure 3.7: Model training and evaluation pipeline. The region of interest (ROI) on the range-velocity image is extracted based on the physical dimensions of the system, the ROI pixel values are flattened into a one-dimensional vector, the models are trained to predict instantaneous mass flow rate and the results are evaluated using root mean square error (RMSE), coefficient of determination (R^2), and symmetric mean average percent error (sMAPE).

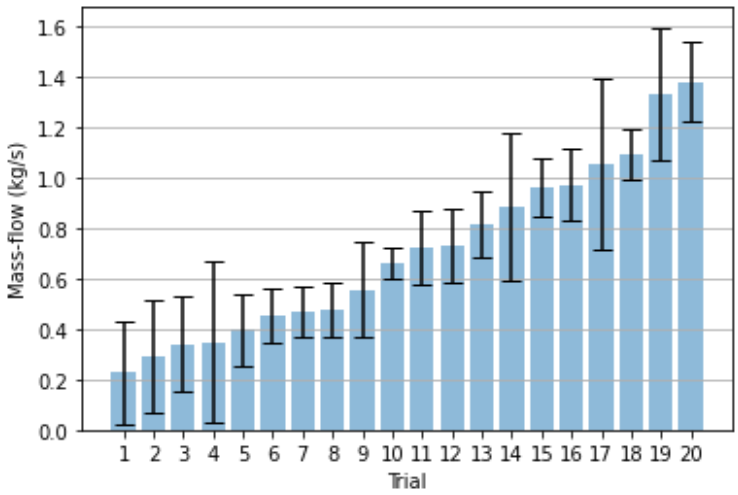


Figure 3.8: Distribution of mass flow rate values in the trials used in the 2-row dataset. The bars represent the mean mass flow rate and the error bars are at one standard deviation of the values distribution.

3.3 Results

3.3.1 Dataset Collection

For the 2-row system, a total of 24 trials were acquired with average mass flows varying from 0.11 to 1.51 kg/s (15 to 200 lb/min). After removing four trials with excessive noise and trimming trials with errors in the beginning and at the end of the acquisition, the dataset consisted of 20 trials that were used for model training and evaluation. Figure 3.8 shows a bar plot with the 20 trials used for the dataset, the bars indicate the average mass flow rate in that trial and the error bars are one standard deviation of the mass flow rate distribution.

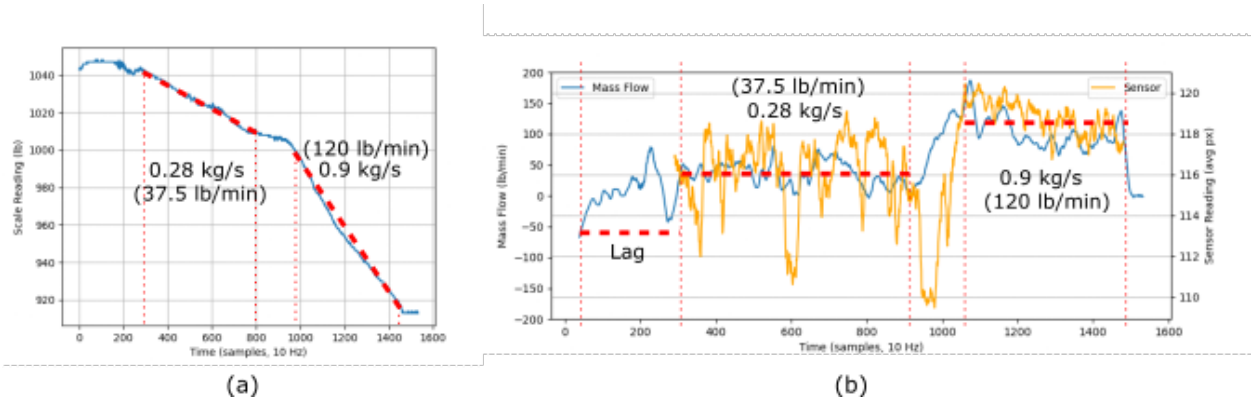


Figure 3.9: Alignment example and magnitude of pixel intensity correlated to mass flow rate for a trial collected with VFD at 70 Hz and gate opening low. (a) Scale measurements for the trial. (b) Mass flow rate calculated using the derivative of the scale and the average pixel value of the radar frame used for alignment.

A sample synchronized trial that was maintained in the dataset is presented in Figure 3.9. Specifically, this is a trial where the VFD was configured at 70 Hz and the gate configuration was for low peanut input. Figure 3.9 shows two dominant mass flow rate periods. The first period was at 0.28 kg/s (37.5 lb/min), at the 100s mark, the peanuts stopped flowing into the air duct and the operator had to intervene. At that moment, the ROI pixel intensity value dropped severely, demonstrating no flow in the duct and after the intervention (i.e., further opening the gate), higher flow of 0.9 kg/s (120 lb/min) was produced. Figure 3.9 also shows that the alignment between the scale-derived mass flow and radar-derived time series, showing that the lag value derived from the alignment process is accurate.

To calibrate the conveyor belt system used to feed peanut mass to the pneumatic conveyor, 11 calibration samples were collected, from 23 to 30 kg (51 to 66 pounds) and with 4.4 to 13.9 seconds resulting in the calibration curve presented in the Figure 3.10c. To build the 6-row combine dataset, 35 trials were acquired for mass flow rate values from 2.3 to 6 kg/s (300 to 800 lb/min). Because the ground truth mass flow rate is determined based on the conveyor calibration model, there is no need to align the two signals temporally. Figure 3.10b shows the mean pixel value and the target mass flow rate for all trials.

	RMSE (kg/s)	**RMSE (lb/min)**	**R^2**	**sMAPE (%)**
Linear Regression	0.19±0.04	25±5	0.75±0.1	24±7
K-Neighbors Regression (n=50)	**0.14±0.04**	**19±5**	**0.85±0.1**	**15±5**
Support Vector Regression (Linear)	0.17±0.04	23±6	0.79±0.13	20±4
Random Forest Regression	0.16±0.02	21±3	0.83±0.11	19±5
Support Vector Regression (RBF)	0.22±0.02	30±3	0.65±0.11	28±4
Multi-Layer Perceptron Regression	0.15±0.05	20±7	0.85±0.11	19±5

Table 3.3: Model comparison results for the 6-row combine dataset. Results are an average of 5-fold cross validation and the 95% confidence interval. Values in bold are overall best results.

3.3.2 Mass Flow Rate Prediction

2-row Combine

Table 3.3 summarizes the results for the fitted models for the 2-row dataset. K-neighbors regressor obtained the lowest RMSE of 0.143 kg/s (19 lb/min) and the lowest sMAPE of 15%. While multi-layer perceptron obtained similar R^2, its RMSE and sMAPE were higher. Figure 3.11 shows scatter plots for the training (Figure 3.11a-e) and test (Figure 3.11f-j) sets for the 5 folds for the best performing model. For the best performing fold (fold 3), the training RMSE is 0.13 kg/s (17 lb/min), R^2 is 0.88 and sMAPE is 16% and the test yielded an RMSE of 0.13 kg/s (17 lb/min),R^2 is 0.89 and sMAPE is 14.35%. Figure 3.11h and 3.11j shows flows noisy ground truth data around 1.1 kg/s (150 lb/min), where the predicted value is constant and the ground truth varies largely.

6-row Combine

Table 3.4 summarizes the results for the fitted models for the 6-row dataset. Random forest regressor obtained the lowest RMSE of 0.52 kg/s (69 lb/min) and the lowest sMAPE of 10%, while k-neighbors regressor obtained the next best RMSE but with worse R^2 and sMAPE. The support vector machine with radial basis function kernel and the multi-layer perceptron models performed the worst and yielded R^2 values worse than random guess. Figure 3.12 shows scatter plots for the training (Figure 3.12a-e) and test (Figure 3.12f-j)

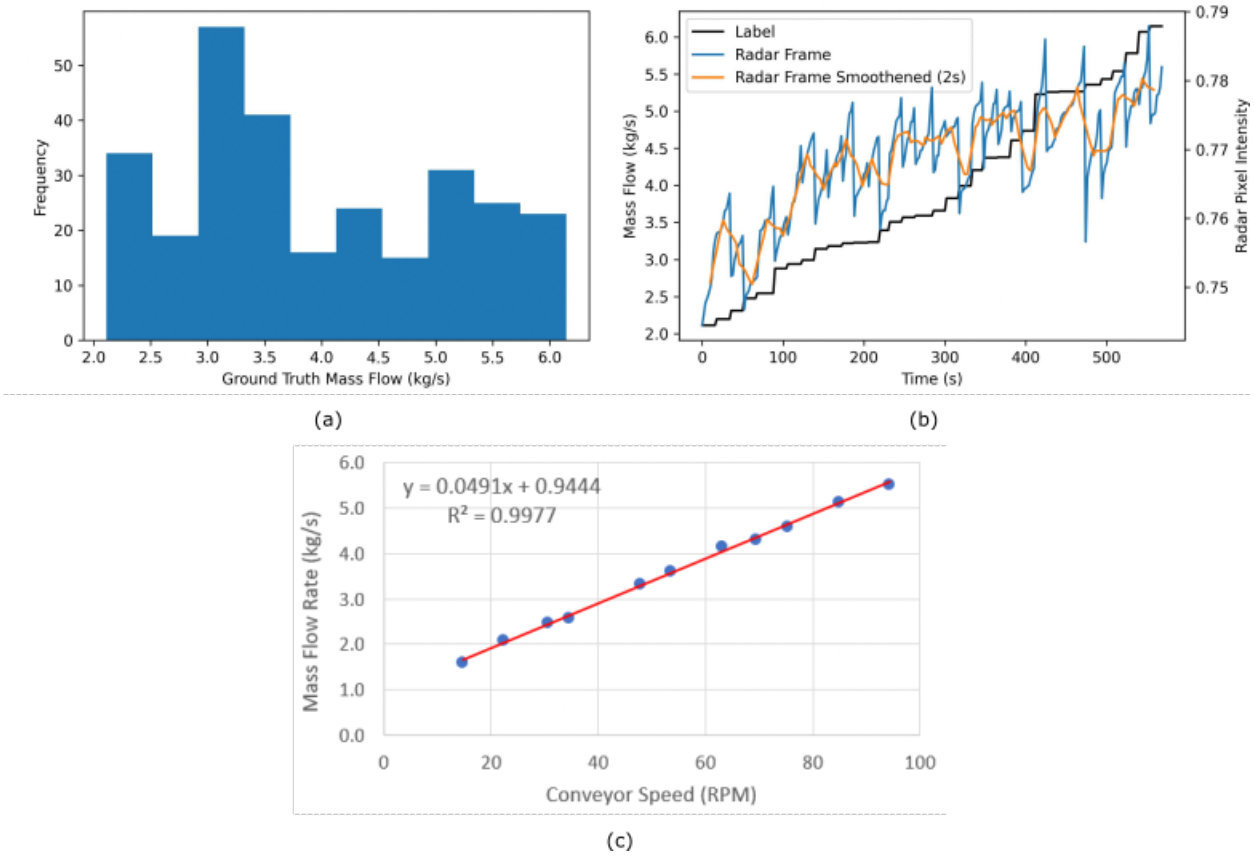


Figure 3.10: 6-row combine dataset visual summary. (a) histogram of ground truth mass flow rate values in the trials for the 6-row combine dataset. (b) sorted average pixel value and the ground truth mass flow rate. (c) conveyor belt calibration rotation-per-minute to peanut mass flow (kg/s) for the 6-row combine system.

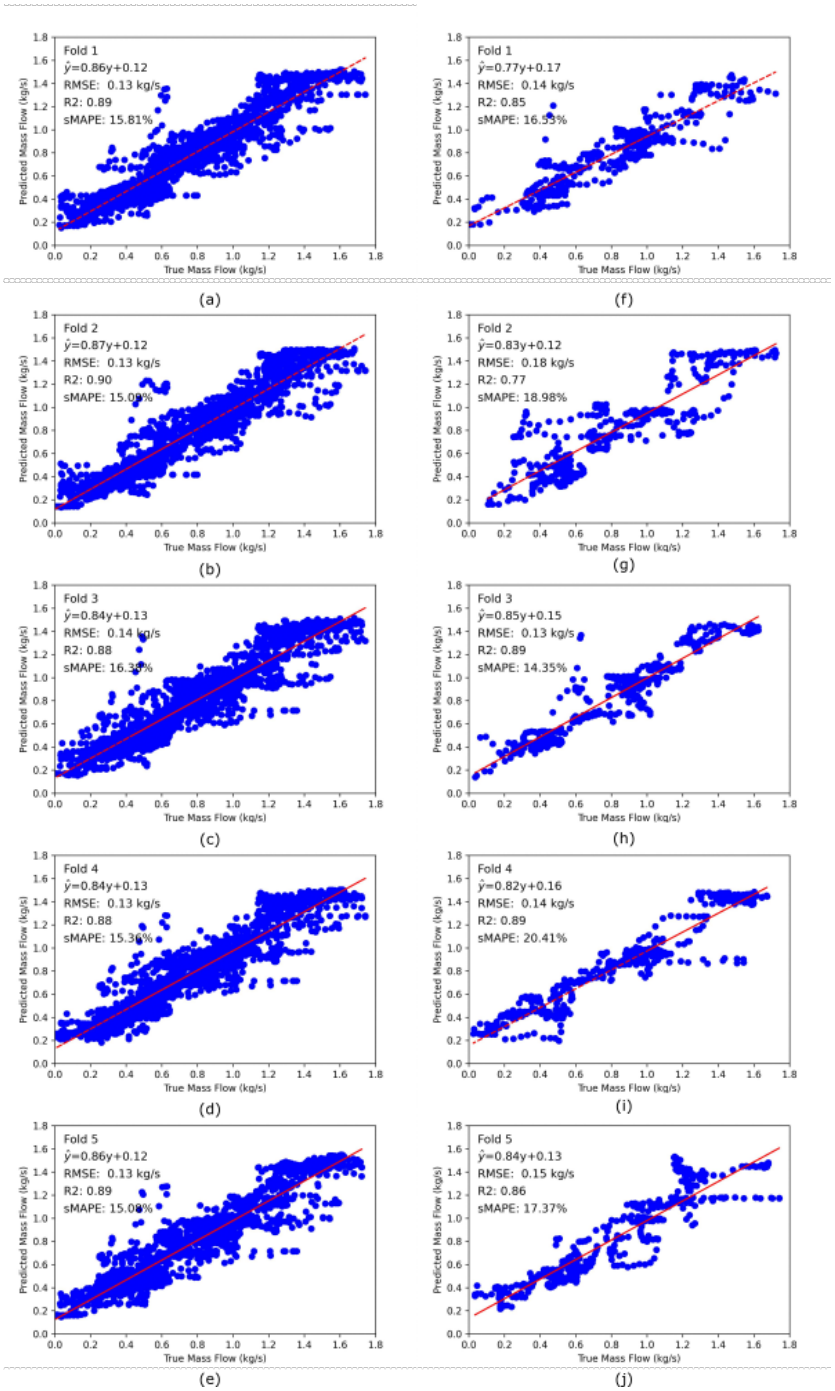


Figure 3.11: Scatter plot from train (a-e) and test (f-j) for 5 folds of k-neighbors regressor in the dataset captured from 2-row combine system. (a) and (f) fold 1, (b) and (g) fold 2, (c) and (h) fold 3, (d) and (i) fold 4, (e) and (j) fold 5.

	RMSE (kg/s)	**RMSE (lb/min)**	**R^2**	**sMAPE (%)**
Linear Regression	0.65±0.17	86±23	0.57±0.44	12±2
K-Neighbors Regressor (n=50)	0.59±0.3	78±40	0.65±0.33	12±4
Support Vector Regression (Linear)	0.65±0.17	86±23	0.57±0.42	12±2
Random Forest Regressor	**0.52±0.26**	**69±35**	**0.71±0.30**	**10±3**
Support Vector Regression (RBF)	1.2±0.79	162±103	-	29±18
Multi-Layer Perceptron Regressor	1.8±1.47	238±195	-	41±36

Table 3.4: Model comparison results for the 6-row combine dataset. Results are an average of 5-fold cross validation at the 95% confidence interval. Values in bold are overall best results.

sets for the 5 folds for the best performing model. For the best performing fold (fold 3), the training RMSE is 0.13 kg/s (17 lb/min), R^2 is 0.88, and sMAPE is 16%. The best performing fold on the test set yielded an RMSE of 0.13 kg/s (17 lb/min), $R^2$is 0.89 and sMAPE is 14.35%.

3.4 Discussion

3.4.1 Data Collection

This work presented two systems that can generate specific mass flow rate values of pneumatically conveyed peanuts for commercial mmWave radar sensors to accurately collect data. Ideally, the system would be configured to produce a constant mass flow so that the sensor can collect accurate data. Though, for lower flow configurations in the 2-row system, the peanut flow would be eventually be interrupted by the peanuts not sliding down the overhead bin. To get the peanuts to flow again, the operator would have to touch the system to open the gate or move the peanuts and force them to flow again. That intervention can be noticed in the variations of the scale readings. One of its effects is that while the peanuts is are flowing, the scale derived mass flow rate results in a negative value. These effects contributed negatively to modeling the mass flow because it added noise into label (scale weight) and therefore the models performed worse with conflicting information in the dataset.

Nevertheless, the different combinations of gate opening and fan speed were able to produce a variety of mass flow values as designed. Figure 3.8 shows that the mean and standard deviation of the scale-derived mass flow varied significantly, but were able to maintain the intended trend to increase and provide a wide range of mass flow rates to train the model and evaluate the sensor. Moreover, Figure 3.9 shows the effectiveness of the proposed alignment algorithm and exemplifies the issues mentioned before about the noise added by the operator. This trial started with a low flow. Then when the flow was interrupted, the operator had to intervene and accidentally interfered with the scale readings. When the operator opened the gate wider, a higher flow was measured.

This noise added by the operator could be avoided with a different design in the data collection system. For instance, an auger could be used to feed peanuts to the duct and the speed of the auger could be controlled to manipulate different mass flow rates. A system using an auger to feed the peanuts would be able to provide controlled mass flow and once calibrated with a scale, the scale would not be necessary during the radar data collection.

The 6-row system implemented a similar strategy to feed peanuts from a calibrated conveyor belt system. Figure 3.10 showed that it was possible to produce a wide range of mass flow rates, but it also shows less increase in the mean pixel intensity near the upper limit of the mass flow range. During the data collection it was noted that when the conveyor belt was running at higher speeds, the mass flow rate was not increasing accordingly due to mechanical issues with the conveyor belt attachment to the combine. Moreover, it was also detected that at a range from 3.8 to 4.9 kg/s (500 to 650 lb/min) the pixel intensities are noisy, which may suggest unstable peanut mass flow rates in that range.

3.4.2 Mass Flow Rate Prediction

Due to the low number of samples in the datasets, it was expected that large deep learning models would not be able to perform well. Also, because of the structure of the data being an image generated from the range-velocity map from a radar system, using

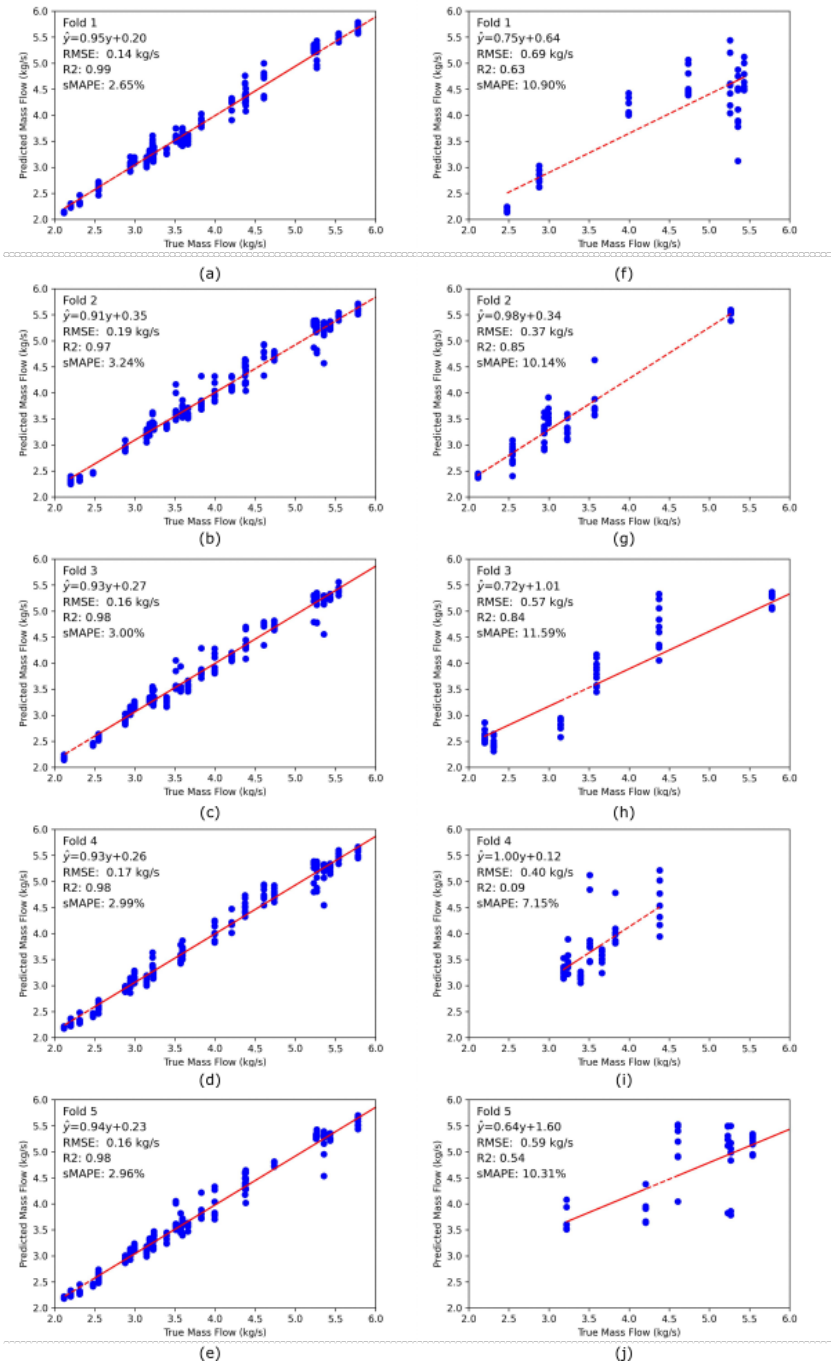


Figure 3.12: Scatter plot from train (a-e) and test (right column) for 5 folds of random forest regressor in the dataset captured from 6-row combine system. (a) and (f) fold 1, (b) and (g) fold 2, (c) and (h) fold 3, (d) and (i) fold 4, (e) and (j) fold 5.

a pre-trained model would not yield a good result. To obtain satisfactory performing in transfer learning, the model has to be pre-trained in a similar task, with similar dataset.

Table 3.3 showed results for the machine learning models for the 2-row dataset. K-neighbors regression achieved the best performance while support vector regression with radial basis function kernel performed the worst. It is known that linear methods like linear regression is sensitive to outliers whereas k-neighbors regression with a large number of neighbors smoothens the effect of noisy labels and outliers. For the 2-row dataset, the test folds 4 and 5 were highly influenced by some predictions around 0.13 kg/s (150 lb/min) but the corresponding labels varied widely. This is yet again another instance of a constant flow inside the duct, but the scale system detected a change in the weight due to an operator interference. Previous studies adapted sensors that were developed for other crops, in this work a sensor was designed to specifically target the harsh environmental conditions of peanut harvesting. Also, the developed sensor predicts instantaneous mass flow rate and can integrate this predicted values to produce a yield estimate.

Table 3.4 showed results for the machine learning models for the 6-row dataset. Random forest regression achieved the best performance while multi-layer perceptron performed the worst. For the 6-row combine dataset, the test folds 4 and 5 had the worst performance, including a R^2 that yielded a worse performance than random guess. This is potentially due to the instable mass flow rate generated at ranges around 3.4 to 4.9 kg/s (450 to 500 lb/min) by the conveyor belt system. An improved design of a peanut feeding system that provides more stable peanut flow cross the entire operating range is expected to provide higher quality calibration data, which subsequently improves mass flow rate prediction.

Thomasson et al., (2006) presented results that are comparable to this study. While evaluating the linearity of their optical sensor for different mass flow rates, a 13 kg bucket was emptied through the sensor in different amounts of duration. Their results show a R^2 of 0.85 when the dumping time was used as reference and a R^2 of 0.95 when the sensor signal timing was used as reference. That indicates that there is noise in the sensor, since the response of

the sensor should be none when no peanut flow was present, which suggests that the sensor accumulated values even when the mass flow was zero. That effect is similarly present in our system, therefore after integrating the mass flow rate, there would be a bias component in our predicted weight. Though, a field-specific calibration can adjust the reading to remove the offset.

Porter et al. (2017) presented calibration curves for yield prediction using a cotton yield monitors for peanut. However, only calibration results were presented and the calibrated model was not further tested with field data. Moreover, the study only presents 9 data points in different fields, which makes it difficult to assess the statistical validity of them. Moreover, they reported problems such as dust buildup and scratches in the sensor lenses and the need for manual compensation of the data. Nevertheless, they presented high-resolution yield maps that could potentially be used for zone management within a precision agriculture approach.

This work has the potential to allow for an in-house pre-calibration that can perform well due to the functioning principle of the sensor that measures the speed and, indirectly, mass of peanuts going through the pneumatic conveyor. Furthermore, better in-field performance can be achieved by combining this study with the system developed by Vellidis et al. (2001). The radar-based sensor can measure mass flow during harvest and when the combine tank is full it can be dumped into a cart with a scale system. The information from the scale can be used to confirm and/or adjust retroactively the values measured by the sensor developed in this study.

Additionally, this system can address another challenge during the peanut harvest related to peanut shelling. When the peanut basket is empty, if the peanut pods hit the back of the cage too fast, the peanut shell can break and release the kernels. That is one of the factors that impact peanut quality at the buying points. Nowadays, the operators visually assess if the peanut pods are reaching the basket and breaking, he/she will adjust a dampener that reduces the air speed in the duct. From the range-velocity image acquired by the

radar, it is possible to extract a measurement of the peanut pod's velocity and allow for better air velocity control. The operator can have a feedback from the radar to adjust to the desirable peanut speed or a control system can be developed to automatically control the damping system based on the velocity signal extracted from the range-velocity image.

3.4.3 Potential Use For Other Applications

This study evaluated the use of an mmWave radar to perform mass flow sensing targeting the development of a peanut yield monitor, however, the sensor can be used for other crops or applications. For instance, to develop yield monitor for other nuts or even measure mass flow of solids in other fields of study (e.g. feed, forestry products, pellets). To diversify the use of such technology to different applications a significant improvement would involve developing mathematical models that can predict the scene in front of the radar based on the sensor reading. The models to compute the mmWave FMCW radar signals, range-FFT, and range-velocity images have already been developed, but the inverse problem has not yet been studied in depth.

3.4.4 Future Work

To effectively evaluate the use of the developed sensor to develop a peanut yield monitor, field tests should be carefully designed. To make sure that the sensor is working in all possible mass flow rates, different ground speeds should be used in similar length rows. Otherwise, the length of the collected sensor would be highly correlated to the total weight and could hide the information from actual field variability relevant for precision crop management. Collecting data with different levels of moisture content is also necessary because of two reasons. First, the water content of the peanuts will give different reflection attenuations in the radio-frequency signal. Second, the weight of the peanuts would change significantly, and therefore the velocity at which they would me travelling through the pneumatic conveyor is also expected to be different. One assumption of the radio systems employed in this work is

that by adjusting the chirp parameters the size of the detectable objects can be manipulated, therefore the sensor should be evaluated for different varieties and soil types to confirm or deny such hypotheses.

Further experiments can be performed by modifying the proportion of the peanuts with foreign materials, moisture content, and peanut varieties so that the model calibrated outside the field transfers well for in-field operation. Yadav et al. (2022) have demonstrated all embedded system for gesture recognition using one of these platforms with good performance and being able to run all radar algorithms and multi-layer perceptron prediction in one standalone system. Similarly, such an approach could be used to produce a compact and weatherproof peanut mass flow sensor that can run the prediction models on board and report directly to the tractor controller. Moreover, for the pre-existing equipment in the field, adding such a system to the combine would only require replacing a segment of the pneumatic conveyor by a plastic section.

3.5 Conclusion

This study developed and evaluated a method for measuring mass flow rate of peanuts using an mmWave FMCW radar-based sensor in two simulated environments of peanut harvesting. One was achieved by retrofitting the pneumatic conveyor of a two-row peanut combine and the other by modifying a 6-row combine into a recirculating system. The proposed machine learning-based model that using k-neighbors regression performed the best for the 2-row harvest combine with an RMSE of 0.14 kg/s (19 lb/min) and a sMAPE of 15% while having a R^2 of 0.85. The best performing model on the 6-row combine dataset was a random forest regressor that achieved RMSE of 0.52 kg/s (69 lb/min), sMAPE of 10% and R^2 of 0.71. The results show potential to overcome the limitations of the harsh environment that was not well addressed by previous research in peanut yield monitor.

Acknowledgements

This project was supported by the Alabama Peanut Producers Association. The authors would like to thank Dan Todd and Hunter Cook of the USDA-ARS National Peanut Research Laboratory personnel for assistance in building the system used in this work and helping with the data collection. As well as James Johnson and Bobby Bradford from the research staff at the Department of Biosystems Engineering at Auburn University for their assistance with building the prototypes.

REFERENCES

Downey, D., Ehsani, R., Giles, K., Haneklaus, S., Karimi, D., Panten, K., Pierce, F., Schnug, E., Slaughter, D., Upadhyaya, S., & Wulfsohn, D. (2010). Advanced Engineering Systems for Specialty Crops: A Review of Precision Agriculture for Water, Chemical, and Nutrient Application, and Yield Monitoring. Landbauforschung, 88.

Henry, D., Aubert, H., & Veronese, T. (2019). Proximal Radar Sensors for Precision Viticulture. IEEE Transactions on Geoscience and Remote Sensing, 57(7), 4624–4635. https://doi.org/10.1109/TGRS.2019.2891886

Hollens Free, D., Kirk, K. R., White, J. W., Brantley, S. A., Peele, J. S., Monfort, W. S., Thomas, J. S., Massey, H. F., Han, Y. J., Montreal, A., & Canada, Q. (2014). Testing of an Impact Plate Yield Monitor for Peanuts: Mounting Configurations and Air Pressure Correction Written for presentation at the 2014 ASABE and CSBE/SCGAB Annual International Meeting Sponsored by.

Kirk, K. R., Porter, W. M., Monfort, W. S., Han, Y. J., Henderson, W. G., & Thomas, J. (2012). Development of a Yield Monitor for Peanut Research Plots. American Society of Agricultural and Biological Engineers Annual International Meeting 2012, ASABE 2012, 3, 1-. https://doi.org/10.13031/2013.41823

Konstantinovic, M., Woeckel, S., Schulze Lammers, P., & Sachs, J. (2008). UWB Radar System for Yield Monitoring of Sugar Beet. Transactions of the ASABE, 51(2), 753–761. https://doi.org/10.13031/2013.24372

Longchamps, L., Tisseyre, B., Taylor, J., Sagoo, L., Momin, A., Fountas, S., Manfrini, L., Ampatzidis, Y., Schueller, J. K., & Khosla, R. (2022). Yield sensing technologies for perennial and annual horticultural crops: a review. Precision Agriculture. https://doi.org/10.1007/s11119-022-09906-2

Pan, E., Tang, J., Kosaka, D., Yao, R., & Gupta, A. (2019). OpenRadar. In GitHub repository. GitHub.

Porter, E., Vellidis, G., Liakos, V., Porter, W., & Branch, B. (2017). An Optical Yield Monitor for Peanuts-Proof of Concept and Evaluation. https://doi.org/10.1017/S2040470017000061

Porter, W. M., Ward, J., Taylor, R. K., & Godsey, C. B. (2020). A Note on the Application of an AgLeader® Cotton Yield Monitor for Measuring Peanut Yield: An Investigation in Two US states. Peanut Science, 47(2), 115–122. https://doi.org/10.3146/PS19-16.1

Rains, G. C., Perry, C. D., & Vellidis, G. (2005). Adaptation and testing of the Agleader cotton yield sensor on a peanut combine. In Applied engineering in agriculture: Vol. v. v. 21.

Thomasson, J. A., Sui, R., Wright, G. C., & Robson, A. J. (2006). Optical Peanut Yield Monitor: Development and Testing. Applied Engineering in Agriculture, 22(6), 809–818.

Vellidis, G., Perry, C. D., Durrence, J. S., Thomas, D. L., Hill, R. W., Kvien, C. K., Hamrita, T. K., & Rains, G. (2001). The Peanut Yield Monitoring System. Transactions of the ASAE, 44(4), 775–785.

Yadav, S. S., Agarwal, R., Bharath, K., Rao, S., & Thakur, C. S. (2022). tinyRadar: mmWave Radar based Human Activity Classification for Edge Computing. 2022 IEEE International Symposium on Circuits and Systems (ISCAS), 2414–2417. https://doi.org/10.1109/ISCAS48785.2022.9937293

Chapter 4

MACHINE VISION-BASED MULTI-ANIMAL DETECTION AND TRACKING FOR BROILER ACTIVITY ANALYSIS

In preparation for Computer and Electronics in Agriculture

Rafael Bidese-Puhl[1], Yin Bao[1], Jeremiah Davis[1], Joseph Purswell[2]

[1] Department of Biosystems Engineering, Auburn University, Auburn, Alabama, USA

[2] Poultry Research Unit, USDA-ARS, Starkville, Mississipi, USA

Highlights

- Development of a comprehensive videos dataset for detect-and-track of broilers.
- Natural lighting proved crucial for more effective model generalization due to diverse data.
- ByteTrack with YOLOv8 outperformed other models in speed and accuracy.

Abstract: *Precision management in poultry farming, utilizing advanced technologies like computer vision and lighting control, has emerged as a promising approach to enhance productivity, meat quality, and animal welfare. These technologies enable effective flock monitoring and data-driven decision making, leading to optimal farm management. Classical models fail when the scene is complex and tracking multiple individual birds is necessary. Detecting-and-tracking allows for an understanding of the spatio-temporal features of the animals positions. This work generated multi-animal detection-and-tracking dataset from top-view videos of broilers at different growth stages and different times of a day. Accuracy and efficiency of different state-of-the-art object detection (YOLOv8 and YOLO-NAS) and trackers (SORT and ByteTrack) were evaluated and compared regarding the generation of a new broiler activity index. From the tracked instances the distance travelled by each bird normalized by the birds' age was used to compose a broiler activity index. Therefore, capturing*

both movement across the pen and compensating for the size of the birds. The results show that the pipeline was able to accurately capture the activity in the pens and quantify activity for animal behavior studies. Further, this pipeline can be extended to determine areas of interest around the pen to identify behaviors in specific regions. Ultimately, once the behaviors are characterized the pipeline can be used to generate feedback signal for control of the broiler houses for light intensity and heater temperatures, providing better conditions of welfare to the animals.

Keywords: Broiler, Detection, MOT, Poultry, Precision Farming, Tracking.

4.1 Introduction

In the context of precision poultry farming, detecting and tracking birds' movements and behaviors can provide valuable insights into their health, welfare, and productivity (Guo et al., 2022). This is achieved through the use of advanced technologies such as computer vision, machine learning, and sensor-based systems (Li et al., 2022; Okinda et al., 2020; van der Sluis et al., 2020). Computer vision, in particular, has emerged as a powerful tool for detecting and tracking birds in a poultry farm setting (Aydin, 2017; Del Valle et al., 2021; Okinda et al., 2020; Zhang et al., 2022). By processing and analyzing video data, computer vision algorithms can identify individual birds, track their movements over time, and even recognize specific behaviors. This allows for continuous, non-invasive monitoring of bird activity, which can be used to assess overall flock health and detect any signs of disease or distress early on. For instance, changes in activity levels or patterns can be indicative of health issues. A sudden decrease in activity might suggest that a bird is sick, while increased aggression could be a sign of stress or overcrowding (Fan et al., 2022; Zukiwsky et al., 2020). By tracking these behaviors, farmers can intervene as necessary to address these issues, improving animal welfare and productivity.

Furthermore, tracking bird activity can also provide insights into their growth and development. For example, tracking the amount of time birds spend eating can help estimate

their feed intake, which is a key factor in their growth rate (Guo et al., 2020; G. Li et al., 2021). Similarly, monitoring the birds' movement can provide information about their physical development and mobility. Machine learning algorithms can also be employed to analyze the collected data and identify patterns or anomalies that might not be immediately apparent. These algorithms can be trained to recognize specific behaviors or predict future outcomes based on historical data, providing a powerful tool for decision-making. Precision management practices in poultry farming have gained significant attention in recent years due to their potential to enhance productivity, improve meat quality, and ensure animal welfare. A key aspect of precision management is the use of advanced technologies, such as computer vision and lighting control, to monitor and manage poultry behavior and health. The use of computer vision in poultry farming has been explored extensively. For instance, the study by Nasiri et al. (2022) demonstrated the potential of 3D computer vision for lameness detection in broiler chickens, a critical aspect of precision management. Similarly, (Guo et al., 2020) developed a chicken distribution analysis system for broiler houses using computer vision, highlighting the technology's potential in monitoring and managing poultry practices.

However, the role of ambient light control as an activity modulator in poultry farming is a relatively new area of research. Light intensity has been shown to influence the behavior and physiological responses of poultry, which can, in turn, impact meat quality. A study by Kuttappan et al., (2016) highlighted the association between increased growth rate in birds and the occurrence of white striping and woody breast myopathies. This suggests that management practices that control growth rate, such as activity modulation through light control, could potentially mitigate these conditions. The woody breast condition, characterized by a tough consistency in raw breast fillets, has been a significant concern in the poultry industry due to its adverse effects on meat quality. Recent studies have started to explore the potential of light control in managing this condition (Mirella Fernandes et al., 2021). For instance, Greene et al., (2019) found that the severity of woody breast myopathy could

be reduced through the modulation of oxygen homeostasis-related genes in broiler chickens, suggesting a potential link between light intensity, activity levels, and the development of woody breast.

To overcome these limitations, several machine vision-based systems have emerged a promising solutions for automating the monitoring and analysis of broiler behavior (Gebhardt-Henrich et al., 2021). In recent years, significant research efforts have been devoted to developing machine vision algorithms and techniques for behavior identification, bird counting (X. Li et al., 2022), lameness detection (Nasiri et al., 2022), and activity level determination (Del Valle et al., 2021; Mirella Fernandes et al., 2021; Neethirajan, 2022). However, conventional approaches often struggle in complex scenes and face challenges when tracking multiple individual birds simultaneously.

To address these issues, this study presents a novel multi-animal detection and tracking pipeline specifically tailored for bird activity monitoring. A comprehensive dataset comprising top-view videos of broilers at different growth stages and various times of the day was generated. This dataset serves as the foundation for evaluating the accuracy and efficiency of state-of-the-art object detection models, including YOLOv8 (Jocher et al., 2023) and YOLO-NAS (Aharon et al., 2021), as well as tracking algorithms such as SORT (Bewley et al., 2016) and ByteTrack (Zhang et al., 2022).

The primary objective of this research is to develop a processing pipeline that is capable of detecting and tracking the birds across different ages to assess their activity levels. The goal is to derive a quantitative measure of broiler activity to optimize feed conversion ratio. Through the analysis of these activity patterns, valuable insights can be gained, contributing to improved management practices that result in better meat quality and productivity.

The specific research objectives were to: (1) generate a comprehensive dataset from top-view videos of broilers at different growth stages and times of the day to create a multi-animal detection-and-tracking dataset for machine vision algorithm training and evaluation; (2) assess the performance of object detection models (e.g. YOLOv8 and YOLO-NAS) to

identify the most accurate and efficient method for detecting individual broilers in complex scenes; (3) evaluate the accuracy and efficiency of SORT and ByteTrack algorithms for tracking multiple broilers simultaneously, identifying the most suitable algorithm for reliable broiler tracking;

4.2 Materials And Methods

4.2.1 Experiment Setup

This experiment was conducted from 06/08/2022 through 07/29/2022 at the Poultry Research Unit of the USDA-ARS in Starkville, MS. In this experiment, 8 pens were used to expose broilers to two different light regimens while other environmental variables were maintained constant (e.g. temperature, humidity, carbon dioxide) to evaluate the occurrence of woody breast. Half of the pens had a side window where the birds were exposed to natural lighting (NL) and half of the pens did not have a window but had intensity-controlled light bulbs installed overhead, these rooms are referred to as artificial lighting (AL). Figure 4.1 shows the different pens, with the feeder, water line and top-view camera installation. During the experiment the light intensity of the AL rooms followed a schedule to expose the birds to darker conditions as they grew older. 42 animals were placed in each room and they had access to food and water at all times. The bedding material in all rooms were the same. These birds were hatched in a nursery and then transported to the growing rooms within 6 hours where they were kept until they were 60 days.

4.2.2 Dataset Collection

High-resolution top-view videos of broilers at different growth stages and various times of the day were recorded using 4MP video cameras, model EKI-K164T412 from Hikvision (Hangzhou, China). The videos aimed to capture the broilers' behaviors and movements within their environment. Two cameras were attached to the ceiling of each growing room, to capture two sides of the room. In total 16 cameras acquired video at a resolution of

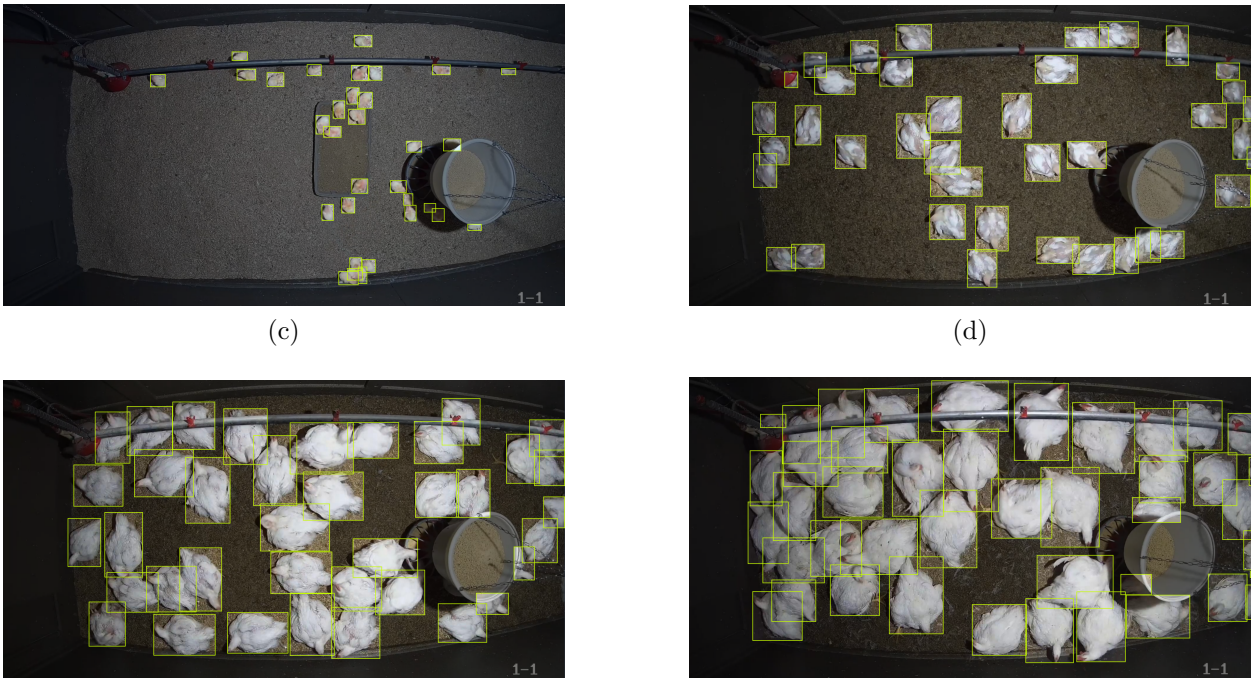
(a) (b)

Figure 4.1: Picture of the pens where the birds were grown. (a) natural lighting and (b) artificial lighting.

1080 × 1920 and at 20 frames-per-second. Due to limitations of storage, a schedule for recording was configured in the acquisition system, the cameras started recording at 6am and stopped recording at 9pm at selected days of the week (Monday, Wednesday, Friday). The video acquisition started in 06/08/2022 and stopped in 07/29/2022. The data was then transferred from the storage system to a workstation for processing.

4.2.3 Data Pre-processing

From the raw data collected different datasets were prepared to perform and evaluate different tasks using videos from four different dates: 06/13/2022, 06/29/2022, 07/15/2022 and 07/29/2022 when the birds were 5, 21, 37 and 51 days old, respectively. From the videos of those four days, selected frames were extracted and annotated to create a dataset to train and evaluate object detection models. Sample frames are shown in Figure 4.2. In total 652 frames were extracted and annotated using the online tool Roboflow (Dwyer et al., 2023) totaling 21934 instances labelled. The birds were annotated using the machine learning automated tool that segments the birds using SAM (Kirillov et al., 2023) and their bounding boxes were extracted based on the bounding boxes of the segmentation masks.



(c) (d)

(e) (f)

Figure 4.2: Sample dataset frames from different bird ages. (a) 06/13/2022; (b) 06/29/2022; (c) 07/15/2022 and (d) 07/29/2022.

Dataset	**Train(Instances)**	**Validation(Instances)**	**Test(Instances)**
AL	325(11091)	287(9498)	40(1345)
NL	327(10843)	285(9632)	40(1459)
NLAL	454(15156)	132(4460)	66(2318)

Table 4.1: Numbers of frames and instances in training, validation and test folds for thedetection datasets AL, NL and NLAL.

The annotated detection dataset was separated into three different datasets for data analysis. First dataset used only frames from AL room for training and frames from the NL room for validation and test. Second, used frames from NL room for training and frames from the AL for validation and test. Finally, a third dataset was composed of a mix of frames from AL and NL, generating the NLAL dataset. For the NLAL dataset, 70% of the frames were randomly chosen to be used for training, the remaining 30% was separated into 10% for validation and 20% for test. Table 4.1 shows the number of frames and instances in each dataset and their respective folds.

Dataset	06/13/2022	06/29/2022	07/15/2022	07/29/2022	Total per camera
AL Camera 1	00:59	01:19	01:09	01:27	04:54
AL Camera 2	01:09	01:09	01:09	01:27	04:54
NL Camera 3	01:17	01:16	01:20	01:19	05:12
NL Camera 4	01:08	01:18	01:28	01:18	05:12
Total per date	04:33	05:02	05:06	05:31	20:12

Table 4.2: Length of the detection datasets AL, NL and NLAL in minutes.

For the multi-object tracking (MOT) task, videos of approximately 1 minute were extracted from the two cameras of one AL room and one NL room at the four bird ages mentioned before. In total, 16 video segments were cut from the original data and annotated using CVAT (Sekachev et al., 2019) annotation tool to produce a MOT dataset for tracking bird position. When annotating the dataset for tracking, the identity of each bird is kept track of from one frame to another, allowing to follow the individual bird throughout the screen. Table 4.2 shows a summary of the durations of the clips and total duration of the dataset.

4.2.4 Object Detection Model Evaluation

Two advanced object detection models, YOLOv8 (Jocher et al., 2023) and YOLO-NAS (Aharon et al., 2021), were selected for evaluation due to their established performance in similar applications. The Ultralytics implementation of the YOLOv8 was compared to the Super-Gradient implementation of YOLO-NAS. The different sizes of the models were compared, for YOLOv8: nano (n), small (s), medium (m), large (l) and extra-large (x); and for YOLO-NAS small (s), medium (m), large (l). The model evaluation was conducted using three different arrangements of the dataset. During training, the models are evaluated using the validation dataset and the training process stops when there is no improvement

in performance after 50 epochs. The hyperparameters for training were kept the same as the default from the libraries. The reported results are mean average precision at 50% intersection-over-union (IoU) (mAP50) and the inference time for each model that were evaluated in unseen test set. The trained models were then tested on unseen test data to assess their accuracy in detecting and localizing individual broilers.

4.2.5 Tracking Algorithm Evaluation

Two tracking algorithms, Simple Online and Realtime Tracking (SORT) (Bewley et al., 2016) and ByteTrack (Zhang et al., 2022) were chosen for evaluating the tracking performance of the broilers. SORT and ByteTrack implementations from the authors were used to evaluate the algorithms. All combinations of detector sizes and trained datasets were evaluated as inputs to the tracking algorithms and their performances were compared using multi-object tracking metrics implemented by py-motmetrics. Most importantly, multi-object tracking accuracy (MOTA) and multi-object tracking precision (MOTP) are evaluated and compared to the execution time of each tracker. MOTA is calculated by Equation 1, where FN is the false negative, FP is false positive, IDSW is the number of id switches and GTDet is the number of ground truth detections. Therefore, the better MOTA performance is the one closer to 100%. This is the metric that better aligns with human visual assessment, but does not include a measure of localization error meaning that the detection performance outweighs the association performance. MOTP is calculated by Equation 2, where TP is the true positive and the second term is the summation of similarity scores over the true positive detections, in this case the similarity metric is Euclidean distance between the two bounding box centers. The best MOTP performance would have the summation term to closer to zero, and therefore the metric would also be closer to 0%. MOTP mostly quantifies the localization accuracy of the detector, and therefore, it provides little information about the actual performance of the tracker.

$$MOTA = \left(1 - \frac{|FN| + |FP| + |IDSW|}{|GTDet|}\right) \cdot 100\% \tag{4.1}$$

$$MOTP = \left(\frac{1}{|TP|}\sum_{TP}\mathcal{S}\right) \cdot 100\% \tag{4.2}$$

4.2.6 Broiler Activity Index Calculation

The tracked instances of each broiler were utilized to measure the distance traveled by each bird across the pen. From one frame to the next, the translation of the centroid of each bird's bounding box is calculated and accumulated to represent the activity of that flock at that moment. The measured distances were normalized by the broilers' age to account for size variations at different growth stages, ensuring a fair comparison of activity levels for birds of different ages. Equation 3 presents this formulation where, F is the total number of frames, f is the current frame, B is the number of bounding boxes in the current frame, b is the current bounding box and c is the center point of the bounding box. The activity metric for the ground truth annotated tracking dataset is used as a reference to compare with the activity index generated by the result of the tracking algorithm. To evaluate if the capacity of each pipeline to produce accurate activity indices that represent the patterns of activity in the pens, the correlation (Equation 4) between the ground truth and predicted activity index (Equation 3). In Equation 4, R is the correlation coefficient matrix and C is the covariance matrix. In this case, R_{01} is used as the point-wise correlation between signals.

$$BAI = \frac{1}{N} \cdot \frac{1}{F}\sum_{f}^{F}\left(\frac{1}{B_f}\sum_{b}^{B_f}\|c_{f-1,b} - c_{f,b}\|_2\right) \tag{4.3}$$

$$R_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}} \tag{4.4}$$

4.2.7 Experiments Environment

The experimental setup employed PyTorch version 1.13.1 with CUDA version 11.7 for the training process. The deep learning modules utilized were Ultralytics version 8.0.110 and Super-Gradients version 3.1.1. The training procedure followed a predefined stopping criterion, whereby training was halted if there was no improvement in performance on the validation set after 50 epochs. On average, the models were trained for 250 epochs. The training was conducted on a high-performance workstation equipped with an AMD Ryzen Threadripper 24-Core Processor, accompanied by 64GB of memory and a NVIDIA Titan RTX GPU with 24GB of memory.

4.3 Results

4.3.1 Object Detection Model Performance

Table 4.3 provides an overview of the evaluation results for various models in terms of their size, frames per second (FPS), and performance metrics. The models under investigation include YOLOv8 and YOLO-NAS, with YOLOv8 further categorized into different sizes (n, s, m, l, x). The size column represents a categorical variable denoting the size of each model, ranging from 'n' to 'x' for YOLOv8 and 's' to 'l' for YOLO-NAS. The parameters column indicates the number of parameters associated with each model, providing insight into their complexity. The FPS column represents the frames per second achieved by each model during inference, reflecting their processing speed. The performance of the models is assessed using four metrics: Average Precision (AP) at 50 IoU (Intersection over Union) thresholds for the three different datasets AL, NL and NLAL. Higher values for these metrics indicate better performance.

Regarding YOLOv8, the 'x' model exhibits the largest number of parameters (XXX). It also achieves the smallest FPS in the YOLOv8 series, indicating slower inference compared to other YOLOv8 models. In terms of performance metrics, the 'x' model achieves the highest

Model	Size	FPS	AL mAP50	NL mAP50	NLAL mAP50
YOLOv8	n	**344**	0.952	0.966	0.988
	s	222	0.965	0.974	0.99
	m	129	0.953	0.983	**0.992**
	l	84	0.975	0.982	**0.992**
	x	57	**0.976**	**0.985**	**0.992**
YOLO-NAS	s	41	0.946	0.952	0.979
	m	34	0.961	0.956	0.981
	l	32	0.962	0.961	0.983

Table 4.3: Detection results using mAP50 for YOLOv8 and YOLO-NAS using AL, NL and NLAL datasets.

scores for the three datasets with mAP50 of 0.952 (AL), 0.985 (NL), and 0.992 (NLAL), implying superior overall performance compared to other YOLOv8 models. Among the all models evaluated, the YOLOv8n model achieved 344 FPS while maintaining a mAP50 scores for AL (0.952), NL (0.966), and NLAL (0.988)

YOLO-NAS, on the other hand, offers models of three different sizes: 's', 'm', and 'l'. The 's' model has the smallest size, while the 'l' model has the largest number of parameters. However, the FPS values for YOLO-NAS models are slightly higher for YOLOv8 models of the same size. In terms of performance, the 'm' model achieves the highest mAP50 scores for AL (0.961), NL (0.956), and NLAL (0.981), still indicating strong performance for object detection tasks.

Similarly, Figure 4.3 shows for the three datasets a plot for mAP performance versus time to inference. Meaning a model in the top-left corner would be good at the object detection task while still being fast to predict its output.

4.3.2 Tracking Algorithm Performance

This section presents the tracking results obtained by evaluating different object detectors in conjunction with two object trackers, ByteTrack and SORT (Table 4.4). The evaluation focused on Multiple Object Tracking Accuracy (MOTA) and Multiple Object

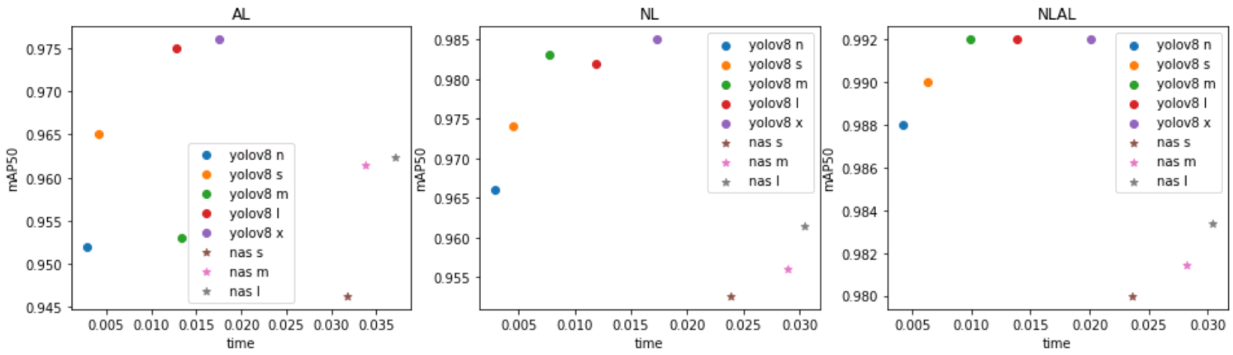


Figure 4.3: Detection results using mAP50 versus prediction time for different sizes of YOLOv8 and YOLO-NAS using AL, NL and NLAL datasets.

Tracking Precision (MOTP) as performance metrics. The trackers were tested on various model sizes and datasets to assess their capabilities.

ByteTrack, paired with the YOLOv8 model, achieved a high tracker frame rate of 99 FPS. It demonstrated consistent performance across different model sizes, ranging from small (s) to extra-large (x). As the model size increased, both MOTA and MOTP scores improved. The largest model, YOLOv8x, achieved the highest MOTA of 79.3 and MOTP of 23.9, operating at 36 FPS. Notably, the best MOTA and MOTP scores were achieved on the NL dataset.

When YOLO-NAS was used with ByteTrack, the frame rate reduced to 66 FPS. However, MOTA performance was lower, with negative values observed for all model sizes. MOTP scores remained relatively high and consistent. SORT as a tracker with the YOLO-NAS model improved MOTA performance for all sizes compared to ByteTrack. However, the best MOTA score achieved was 4 for the YOLO-NAS(m) configuration. The best MOTP of 22.7 was achieved by YOLO-NAS(l) with SORT. Once again, the best MOTA and MOTP performances were observed on the NL dataset. SORT, in combination with the YOLOv8 models as object detectors, achieved a high tracker frame rate of 205 FPS. Among the different model sizes, YOLOv8L performed the best with an MOTA score of 80.2 at an FPS of 57, and a MOTP of 23.7. The fastest model, YOLOv8n, maintained a MOTA of 61.2 and

Tracker	Model	Tracker FPS	Size	Total FPS	AL MOTA ↑	AL MOTP ↓	NL MOTA ↑	NL MOTP ↓	NLAL MOTA ↑	NLAL MOTP ↓
ByteTrack	YOLOv8	99	n	74	59.7	36.4	67.2	25.3	59.1	30.6
			s	66	65.3	35.6	73.8	24.7	66.6	30.3
			m	50	67.3	35.5	74.4	24.4	68.5	30.2
			l	44	64.9	36.1	79.0	24.0	70.6	29.9
			x	36	63.7	36.0	79.3	23.9	72.4	29.7
	NAS	66	s	24	-7.6	36.4	-1.3	26.7	-2.4	30.1
			m	22	-6.5	36.9	-2.5	26.9	-4.0	31.4
			l	21	-5.9	36.9	-3.9	26.8	-4.6	31.8
SORT	YOLOv8	**205**	n	**123**	61.2	36.5	61.2	25.2	60.0	30.5
			s	102	66.9	35.7	73.9	24.3	67.9	30.2
			m	68	68.5	35.5	77.0	24.1	69.5	30.1
			l	57	66.2	36.2	**80.2**	23.7	71.5	29.8
			x	43	64.9	36.1	79.3	23.6	73.3	29.6
	NAS	138	s	30	3.0	34.7	3.5	23.8	4.0	26.8
			m	27	3.0	35.7	4.0	23.1	3.9	28.6
			l	25	3.8	35.9	3.2	**22.7**	3.6	29.0

Table 4.4: Tracking results using MOTA and MOTP for YOLOv8 and YOLO-NAS using AL, NL and NLAL datasets for detector training.

MOTP of 25.2. The NL dataset showcased the best MOTA and MOTP performances for SORT.

The graph illustrates the performance comparison of object trackers, ByteTrack and SORT, using different models, namely YOLOv8 and YOLO-NAS, on three datasets: AL, NL, and NLAL. The x-axis represents the time taken for prediction, while the y-axis represents the MOTA (Figure 4.4) and MOTP (Figure 4.5) scores.

4.3.3 Broiler Activity Index Calculation

Table 4.5 provides a summary of the tracking performance correlation results for different combinations of object trackers (ByteTrack and SORT) and object detection models (YOLOv8 and YOLO-NAS) on three datasets (AL, NL, and NLAL). The table includes

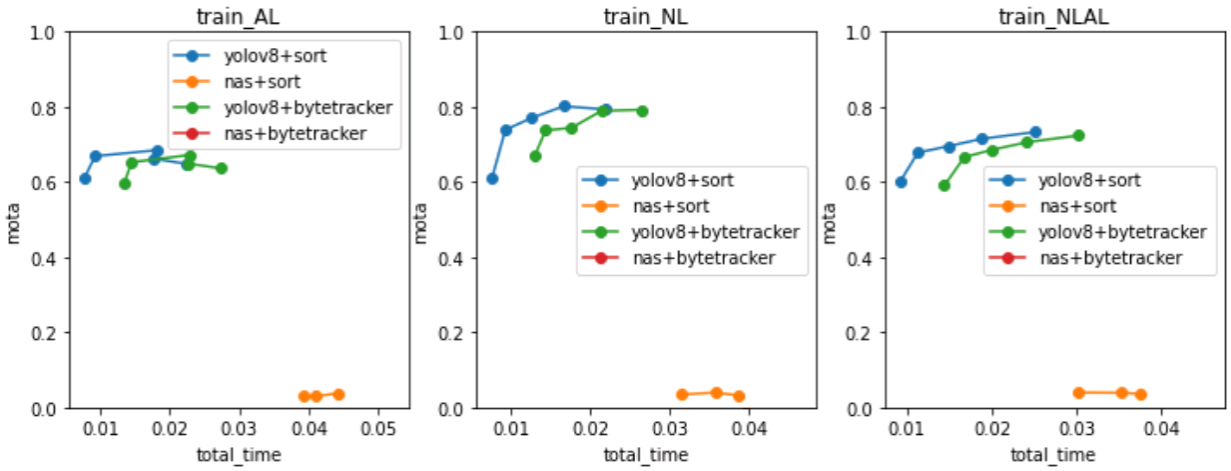


Figure 4.4: Tracking results showing MOTA versus prediction time for different sizes detectors trained on AL, NL and NLAL datasets.

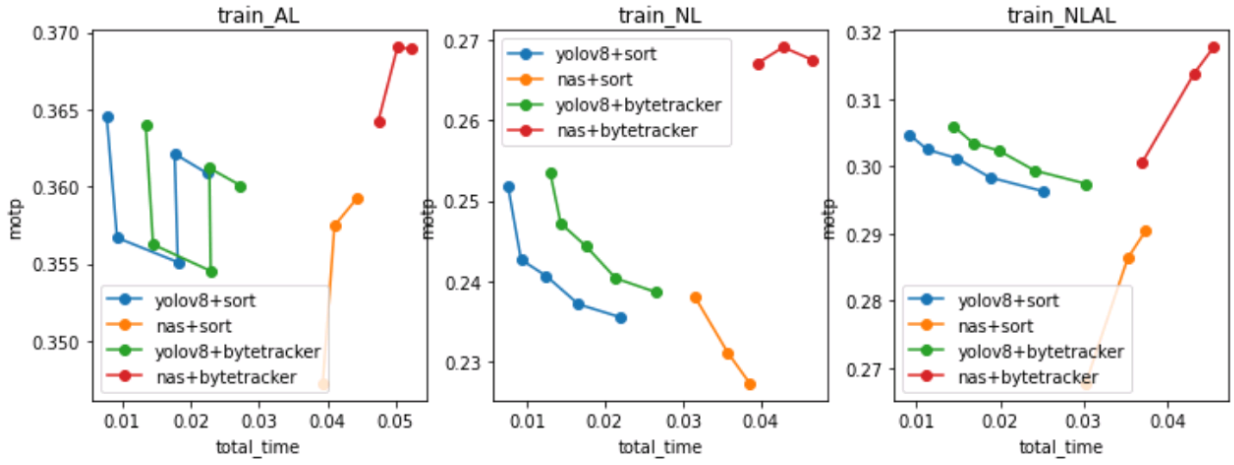


Figure 4.5: Tracking results showing MOTP versus prediction time for different sizes detectors trained on AL, NL and NLAL datasets.

Tracker	Model	Size	AL Corr	NL Corr	NLAL Corr
ByteTrack	YOLOv8	n	**96.5**	95.8	**97.1**
		s	94.0	**97.0**	96.4
		m	90.6	87.7	93.3
		l	89.2	89.5	92.8
		x	87.8	85.9	91.9
	YOLO-NAS	s	57.7	43.3	87.3
		m	32.9	96.6	19.1
		l	82.4	54.5	93.7
SORT	YOLOv8	n	93.7	91.7	96.4
		s	93.0	95.6	94.2
		m	89.4	88.5	90.3
		l	88.7	86.1	89.7
		x	85.3	85.4	87.4
	YOLO-NAS	s	93.5	62.5	61.0
		m	60.7	51.0	78.8
		l	58.0	77.9	4.2

Table 4.5: Activity index correlation between ground truth tracking data and predicted with different trackers and detectors trained on AL, NL and NLAL datasets.

correlation values that indicate the accuracy of the trackers in predicting the positions of tracked objects.

For each tracker-model combination, the table presents the correlation values for various model sizes (n, s, m, l, x) on each dataset. Higher correlation values indicate better alignment between the predicted positions and the actual positions of the objects being tracked. ByteTrack with YOLOv8 achieved high correlation scores across all model sizes and datasets, consistently demonstrating accurate tracking performance. On the other hand, ByteTrack with YOLO-NAS exhibited varying correlations, with some sizes performing well on certain datasets and poorly on others. Similarly, SORT with YOLOv8 displayed strong correlations, while SORT with YOLO-NAS had mixed results across different sizes and datasets.

Figure 4.6 presents the correlation between predicted and ground truth object positions for three datasets: AL, NL, and NLAL. The x-axis denotes the size of the models, ranging from small (s) to extra-large (x), including YOLOv8 models (n, s, m, l, x) and YOLO-NAS

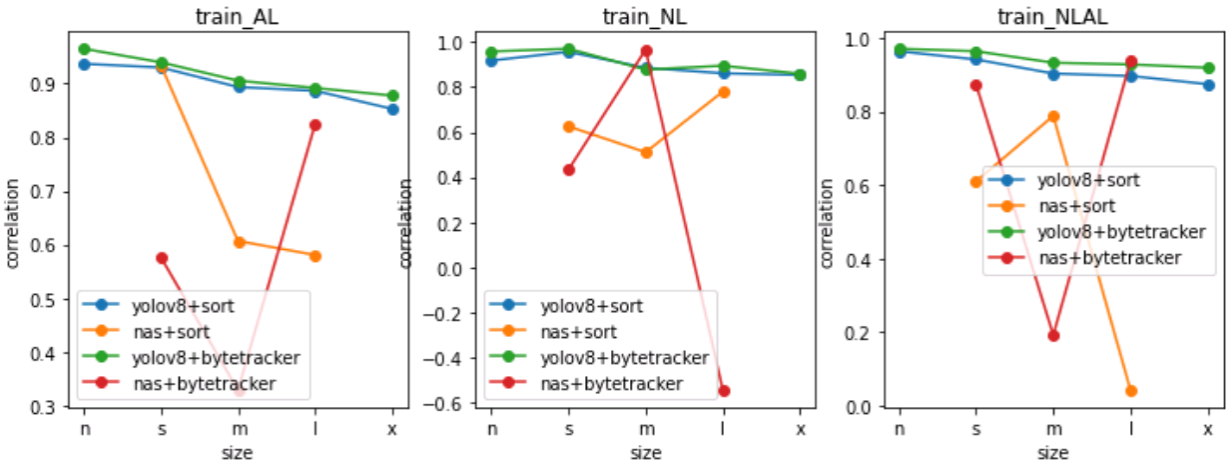


Figure 4.6: Activity index correlation versus detector size for different detectors trained on AL, NL and NLAL datasets.

models (s, m, l). The y-axis represents the correlation values, indicating the accuracy of the trackers in predicting object positions.

4.4 Discussion

In this study, an expansive dataset tailored for broiler detection and multi-animal tracking was generated. This dataset provides a robust foundation for training machine learning models with the aim of effectively identifying broilers and tracking multiple animals. The results of this study demonstrated the effectiveness of the developed pipeline in accurately capturing the activity of broilers within the pens and quantifying their activity levels. The object detection models, YOLOv8 and YOLO-NAS, showcased high accuracy in detecting and localizing individual broilers, even in complex scenes. This finding highlights the potential of these models in accurately identifying broiler behaviors and facilitating subsequent analysis.

Furthermore, comprehensive evaluation of the tracking algorithms utilized in this study. For this purpose, we employed two established tracking metrics: Multiple Object Tracking Accuracy (MOTA) and Multiple Object Tracking Precision (MOTP). These metrics provide

quantitative measures of the tracking performance, facilitating a clear and objective assessment of the developed algorithms. The evaluation process, thereby, ensures that the quality, efficiency, and robustness of our algorithms are meticulously assessed and validated.

Despite being smaller in size and having less parameters compared to the 's', 'm', 'l', and 'x' models, the 'n' model achieves a remarkable FPS value. This implies that the 'n' model offers an efficient inference speed, making it suitable for real-time applications. Furthermore, the 'n' model maintains a high mAP50, indicating accurate object detection capabilities. The impressive combination of superior FPS performance and high mAP50 makes the 'n' model a compelling choice for scenarios that demand both speed and accuracy in object detection tasks.

Interestingly, the models trained on the NL dataset exhibited a superior performance compared to the others. The enhanced performance can be attributed to the inherent diversity present in the NL dataset. This dataset encapsulates a broader range of lighting conditions compared to the AL dataset, as it includes variations arising from light intensity, shadows and other artifacts resulting from the white balance and automatic compensations from the camera. This variety enriches the dataset, thus presenting a more comprehensive and challenging learning environment for the model. Consequently, this trains the model to better generalize and adapt to novel lighting conditions, leading to superior performance when compared to the models trained solely on artificial lighting or a mix of both.

The evaluation of object detection models and tracking algorithms was based on the specific dataset and experimental conditions used. The generalizability of the results to different broiler populations and settings should be considered carefully. Future research could focus on expanding the dataset and considering additional variables that may influence broiler behaviors, such as environmental factors or group dynamics.

4.4.1 Model Performance

Reviewing the Table 4.3, YOLOv8 consistently performs with high speed (FPS) and accuracy (mAP50) across all sizes and lighting conditions. It reaches peak FPS at the nano size (344 FPS) and maintains high mAP50 scores in all scenarios, peaking at 0.992 under mixed lighting conditions (NLAL). The mAP50 score also seems to generally improve as the size parameter increases. YOLO-NAS, on the other hand, runs at a significantly slower speed compared to YOLOv8. Its performance in terms of mAP50 is also generally lower than that of YOLOv8, but it still performs reasonably well with mAP50 scores all above 0.94. Interestingly, for YOLO-NAS, the mAP50 scores increase slightly as the size parameter increases from small to large. In summary, YOLOv8 outperforms YOLO-NAS in terms of both speed and accuracy across all size parameters and lighting conditions. However, YOLO-NAS still demonstrates decent performance, especially considering the mAP50 values under the NLAL condition.

Looking at the MOTA scores (where a higher score is better), the ByteTrack model with YOLOv8 detection performed consistently well under all lighting conditions and sizes. The performance was especially good under natural lighting conditions, with MOTA scores ranging from 67.2 to 79.3. On the other hand, the YOLO-NAS model under ByteTrack showed negative MOTA scores indicating poor performance, particularly under artificial light conditions. The SORT model with YOLOv8 also showed good performance, with MOTA scores slightly lower than ByteTrack under similar conditions but significantly better than the ByteTrack with YOLO-NAS.

In terms of MOTP (where a lower score is better), all models showed similar performance. There was no clear winner, indicating that all models and combinations were about equally precise in their tracking. In terms of speed, SORT with YOLOv8 demonstrated the highest Tracker FPS, particularly in the nano size setting. This implies that it might be the best choice for applications where speed is a critical factor. In summary, the ByteTrack model with YOLOv8 outperforms others in tracking accuracy across all lighting conditions

and sizes, whereas the SORT model with YOLOv8 has the advantage in terms of processing speed.

4.4.2 Broiler Activity Index

The correlation score here refers to the degree to which the predictions of the models align with the true values, with a higher score indicating a stronger correlation, and thus, better model performance. From the results, ByteTrack with YOLOv8 consistently demonstrates strong correlations across all sizes and lighting conditions, with scores generally above 85%. The highest correlation score is observed under the NLAL lighting condition at nano size, with a score of 97.1.

The performance of ByteTrack with YOLO-NAS is more varied, with some sizes showing low correlation scores under certain conditions. For instance, under the NL lighting condition at medium size, there is a high correlation score of 96.6, but under the NLAL lighting condition, the correlation score drops dramatically to 19.1.

The SORT model with YOLOv8 shows strong and consistent correlations across all sizes and lighting conditions as well, but generally slightly lower than ByteTrack with YOLOv8. SORT with YOLO-NAS, however, demonstrates more inconsistent results, with correlation scores ranging widely from a low of 4.2 (large size under NLAL condition) to a high of 93.5 (small size under AL condition).

In summary, the ByteTrack model with YOLOv8 consistently delivers the strongest correlation scores across different sizes and lighting conditions. This suggests that its predictions align very well with the true values, and it is thus likely the most reliable model among those tested here. The performance of the models using YOLO-NAS is more variable, with some combinations performing quite well and others poorly, suggesting that the success of this model may be more dependent on the specific conditions or in better fine tuning of the model.

4.4.3 Future Work

While YOLOv8, in conjunction with ByteTrack, demonstrates superior performance under our current conditions, there is a clear need to explore how we can improve the performance of YOLO-NAS, given its inconsistent results across different conditions and sizes. A possible approach could involve fine-tuning the YOLO-NAS model or modifying its architecture to better handle variances in lighting conditions and object sizes.

In addition to the primary objectives, this study also identified potential extensions and applications of the developed pipeline. The ability to determine areas of interest within the pen opens up possibilities for identifying specific behaviors in different regions. This feature enhances the pipeline's versatility and enables focused analysis of broiler behaviors based on spatial considerations. Moreover, the integration of the pipeline with broiler house control systems holds promise for improving meat quality and feed conversion ratio. By utilizing the generated feedback signals, adjustments to factors such as light intensity and heater temperatures can be made, creating a more favorable environment for the birds.

4.5 Conclusion

In this study, a comprehensive exploration of broiler detection and multi-animal tracking using state-of-the-art models was conducted. The use of the ByteTrack and SORT models, in conjunction with YOLOv8 and YOLO-NAS, allowed to investigate the performance of these models under different lighting conditions and object sizes. The findings show that ByteTrack, when coupled with YOLOv8, consistently demonstrates high performance across various conditions and sizes, both in terms of speed and accuracy. On the other hand, while YOLO-NAS showed potential, its inconsistent performance indicates that further optimization may be necessary for it to be as effective across a broad range of scenarios.

It was found that natural lighting conditions provide a more diverse training environment that results in superior model generalization, underscoring the importance of having varied and robust datasets for training machine learning models.

Despite the progress made, this study also highlights potential avenues for future research. While the performance of these models was promising, there is room for further optimization, especially in terms of processing speed for real-time applications and model robustness under diverse environmental conditions.

Acknowledgments

The authors would like to thank the Poultry Research Unit of the USDA-ARS and its personnel for their invaluable support and collaboration throughout this research project. Their contributions have not only facilitated this research but have also been pivotal in the progress and success of this study.

REFERENCES

Aharon, S., Louis-Dupont, Ofri Masad, Yurkova, K., Lotem Fridman, Lkdci, Khvedchenya, E., Rubin, R., Bagrov, N., Tymchenko, B., Keren, T., Zhilko, A., & Eran-Deci. (2021). Super-Gradients. In GitHub repository. GitHub. https://doi.org/10.5281/ZENODO.7789328

Aydin, A. (2017). Using 3D vision camera system to automatically assess the level of inactivity in broiler chickens. Computers and Electronics in Agriculture, 135, 4–10. https://doi.org/10.1016/J.COMPAG.2017.01.024

Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016). Simple online and realtime tracking. 2016 IEEE International Conference on Image Processing (ICIP), 3464–3468. https://doi.org/10.1109/ICIP.2016.7533003

Del Valle, J. E., Pereira, D. F., Mollo Neto, M., Gabriel Filho, L. R. A., & Salgado, D. D. A. (2021). Unrest index for estimating thermal comfort of poultry birds (Gallus gallus domesticus) using computer vision techniques. Biosystems Engineering, 206, 123–134. https://doi.org/10.1016/J.BIOSYSTEMSENG.2021.03.018

Dwyer, B., Nelson, J., & Solawetz, J. (2023). Roboflow (No. 1). Roboflow. Fan, B., Bryant, R., & Greer, A. (2022). Behavioral Fingerprinting: Acceleration Sensors for Identifying Changes in Livestock Health. J 2022, Vol. 5, Pages 435-454, 5(4), 435–454. https://doi.org/10.3390/J5040030

Gebhardt-Henrich, S. G., Stratmann, A., & Dawkins, M. S. (2021). Groups and individuals: Optical flow patterns of broiler chicken flocks are correlated with the behavior of individual birds. Animals, 11(2), 1–9. https://doi.org/10.3390/ANI11020568

Greene, E., Flees, J., Dadgar, S., Mallmann, B., Orlowski, S., Dhamad, A., Rochell, S., Kidd, M., Laurendon, C., Whitfield, H., Brearley, C., Rajaram, N., Walk, C., & Dridi, S. (2019). Quantum Blue Reduces the Severity of Woody Breast Myopathy via Modulation of Oxygen Homeostasis-Related Genes in Broiler Chickens. Frontiers in Physiology, 10, 460105. https://doi.org/10.3389/FPHYS.2019.01251/BIBTEX

Guo, Y., Aggrey, S. E., Wang, P., Oladeinde, A., & Chai, L. (2022). Monitoring Behaviors of Broiler Chickens at Different Ages with Deep Learning. Animals 2022, Vol. 12, Page 3390, 12(23), 3390. https://doi.org/10.3390/ANI12233390

Guo, Y., Chai, L., Aggrey, S. E., Oladeinde, A., Johnson, J., & Zock, G. (2020). A Machine Vision-Based Method for Monitoring Broiler Chicken Floor Distribution. Sensors, 20(11), 3179. https://doi.org/10.3390/s20113179

Jocher, G., Chaurasia, A., & Qiu, J. (2023). YOLO by Ultralytics. https://github.com/ultralytics/ultralytics

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., & Girshick, R. (2023). Segment Anything. https://arxiv.org/abs/2304.02643v1

Kuttappan, V. A., Hargis, B. M., & Owens, C. M. (2016). White striping and woody breast myopathies in the modern poultry industry: a review. Poultry Science, 95(11), 2724–2733. https://doi.org/10.3382/PS/PEW216

Li, G., Hui, X., Chen, Z., Chesser, G. D., & Zhao, Y. (2021). Development and evaluation of a method to detect broilers continuously walking around feeder as an indication of restricted feeding behaviors. Computers and Electronics in Agriculture, 181. https://doi.org/10.1016/J.COMPAG.2020.105982

Li, X., Zhao, Z., Wu, J., Huang, Y., Wen, J., Sun, S., Xie, H., Sun, J., & Gao, Y. (2022). Y-BGD: Broiler counting based on multi-object tracking. Computers and Electronics in Agriculture, 202. https://doi.org/10.1016/J.COMPAG.2022.107347

Mirella Fernandes, A., de Lucca Sartori, D., José de Oliveira Morais, F., Salgado, A., Florentino Pereira, D., Oliveira Morais, de, Analysis, D., Cambra-López, M., Jorge De Moura, D., & Zheng, W. (2021). Analysis of Cluster and Unrest Behaviors of Laying Hens Housed under Different Thermal Conditions and Light Wave Length. Animals 2021, Vol. 11, Page 2017, 11(7), 2017. https://doi.org/10.3390/ANI11072017

Nasiri, A., Yoder, J., Zhao, Y., Hawkins, S., Prado, M., & Gan, H. (2022). Pose estimation-based lameness recognition in broiler using CNN-LSTM network. Computers and Electronics in Agriculture, 197. https://doi.org/10.1016/J.COMPAG.2022.106931

Neethirajan, S. (2022). ChickTrack – A quantitative tracking tool for measuring chicken activity. Measurement: Journal of the International Measurement Confederation, 191. https://doi.org/10.1016/J.MEASUREMENT.2022.110819

Okinda, C., Nyalala, I., Korohou, T., Okinda, C., Wang, J., Achieng, T., Wamalwa, P., Mang, T., & Shen, M. (2020). A review on computer vision systems in monitoring of poultry: A welfare perspective. Artificial Intelligence in Agriculture, 4, 184–208. https://doi.org/10.1016/J.AIIA.2020.09.002

Sekachev, B., Manovich, N., & Zhavoronkov, A. (2019). Computer Vision Annotation Tool. https://doi.org/10.5281/ZENODO.3497106

van der Sluis, M., de Haas, Y., de Klerk, B., Bas Rodenburg, T., & Ellen, E. D. (2020). Assessing the Activity of Individual Group-Housed Broilers Throughout Life Using

a Passive Radio Frequency Identification System—A Validation Study. Sensors 2020, Vol. 20, Page 3612, 20(13), 3612. https://doi.org/10.3390/S20133612

Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2022). ByteTrack: Multi-Object Tracking by Associating Every Detection Box. Proceedings of the European Conference on Computer Vision (ECCV).

Zukiwsky, N. M., Girard, T. E., & Zuidhof, M. J. (2020). Effect of an automated marking system on aggressive behavior of precision-fed broiler breeder chicks. Journal of Applied Poultry Research, 29(4), 786–797. https://doi.org/10.1016/J.JAPR.2020.06.005