

COMPUTER VISION APPLIED TO BEEF CATTLE PRODUCTION

by

Joseane Padilha da Silva

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy
(Animal and Dairy Sciences)

at the

UNIVERSITY OF WISCONSIN-MADISON

2023

Date of final oral examination: 12/01/2023

The dissertation is approved by the following members of the Final Examination Committee:

Guilherme J. M. Rosa, Professor, Animal and Dairy Sciences

Joao R. R. Dórea, Assistant Professor, Animal and Dairy Sciences

Francisco Peñagaricano, Assistant Professor, Animal and Dairy Sciences

William A. Sethares, Professor, Electrical Engineering

© Copyright by Joseane Padilha da Silva 2023

ALL RIGHTS RESERVED

DEDICATION

I dedicate this dissertation to three incredible individuals who have been the foundation, inspiration, and driving force behind every step of this academic and personal journey.

To my beloved spouse, Odille, who has been my unwavering support, my endless source of love, and my reason to persevere. Your love, support and understanding are the light that illuminates every page of this work.

To my dear mother, Maria, whose love, guidance, and strength have shaped the woman I am today. Your tireless dedication to my education taught me the true meaning of perseverance.

To my father, Gilberto, whose sweetness, humility, honesty, and inspiring example remain a constant source of strength and pride for me.

To this family, both the one I was born into and the one I've built with my wife (extended to Lego and Rini), I thank you for being my anchor and motivation. This work is an expression of the love and gratitude I feel for each of you.

ACKNOWLEDGMENTS

I would like to begin by expressing my gratitude to Dr. Guilherme Rosa, my advisor, for helping me become a researcher, for making me believe in my potential, for turning my professional dreams into reality, and for being so generous in imparting your knowledge effectively, which greatly assisted in my PhD.

Thank you to Dr. Bill Sethares for his invaluable guidance on my examination committee, including his lessons in image processing. I also want to express my gratitude to Dr. Joao Dorea and Dr. Francisco Peñagaricano for their dedicated review of my materials and valuable insights.

Gratitude also goes to all the staff at the Department of Animal and Dairy Sciences, particularly Debby Schneider, Megan Sippel, Dan Mechenich, Minh Ngo, and Steve Switzer, for their valuable support. It is even more remarkable that they assist us with such joy and dedication. Additionally, my appreciation to my colleagues Arthur Fernandes, Anderson Alves, and Rafael Ferreira for their contributions to my thesis.

I'd like to give a special thanks to my office colleagues, Fiona Guinan and Negin Sheybani. Your company has been a source of joy, and our friendship extends beyond our academic journey.

I am profoundly thankful to EMBRAPA for enabling my studies. My goal is to serve my country responsibly and passionately in agricultural research to contribute to its progress.

A huge thank you also to my nieces and nephew Emille, Maria Clara and Joao Martin. To my beloved brother, Ricardo.

Finally, I want to express my gratitude to an angel in heaven who couldn't witness my doctoral achievement but continues to guide and support me. To grandmother Florinda, my heartfelt thanks for your inspiring wisdom.

ABSTRACT

The global demand for high-quality protein has placed the beef industry in a pivotal role. This dissertation explores the application of computer vision to address critical challenges within the beef cattle production. Chapter one provides a general overview of the topic and an outline for the contents of the dissertation. Chapter two presents a comprehensive pipeline tailored to process top-view images of beef cattle in outdoor feedlot environments. This pipeline aims to extract vital features for subsequent body weight predictions, tackling technical hurdles such as varying light conditions, automatic single-animal image classification, strategies for enhancing classifier performance using data augmentation with synthetic images, and automatic body-segmentation. Chapter three focus on the evaluation of two feature extraction methods: traditional image-derived features and Fourier transformation. This chapter demonstrates the superiority of Fourier transformation, especially when combined with artificial neural networks, for body weight prediction. Chapter four explores computer vision algorithms for animal identification using facial images across locations in the meat production chain. The findings provide critical insights into advancing image processing techniques and exploring alternative model architectures for image-based animal identification and tracking. Lastly, a concluding chapter provides some closing remarks, underscoring that this research offers a glimpse into the future of animal husbandry, emphasizing automation while minimizing stress on animals and human intervention. It also highlights the importance of image processing techniques, especially when they are used in conjunction with machine learning methods to handle images, especially those collected in challenging real-world conditions.

TABLE OF CONTENTS

| | |
|---|------|
| DEDICATION | i |
| ACKNOWLEDGMENTS | ii |
| ABSTRACT..... | iii |
| TABLE OF CONTENTS..... | iv |
| LIST OF TABLES | vii |
| LIST OF FIGURES | viii |
| CHAPTER ONE: INTRODUCTION..... | 1 |
| References..... | 5 |
| CHAPTER TWO: A PIPELINE APPROACH FOR 3D IMAGE PRE-PROCESSING, CLASSIFICATION, SEGMENTATION, AND FEATURE EXTRACTION FOR FUTURE WEIGHT PREDICTION OF BEEF CATTLE IN FEEDLOTS..... | 8 |
| Abstract | 8 |
| Introduction | 9 |
| Material and Methods | 13 |
| Data acquisition_ | 13 |
| Image processing and Computer Vision pipeline | 13 |
| First Module: Image Pre-Processing | 15 |
| Second Module: Image classification | 16 |
| AlexNet | 17 |
| ResNet-50 with support vector machine | 18 |
| Third Module: Image Augmentation with Synthetic Images | 18 |
| Generative Adversarial Networks | 18 |
| Generating synthetic images using GANs | 20 |

| | |
|---|-----------|
| Classifier Selection | 21 |
| Fourth Module: Animal Body’s Segmentation | 22 |
| Image type comparison for semantic segmentation | 22 |
| Pos-processing on segmented images | 23 |
| Marker-Controlled Watershed Segmentation (MCWS) | 24 |
| Classifier-based filtering and false positive reduction | 25 |
| Feature extraction | 26 |
| Results and Discussion | 27 |
| First Module: Image Pre-Processing | 27 |
| Second Module: Image classification | 27 |
| Third Module: Image Augmentation with Synthetic Images | 28 |
| Fourth Module: Animal Body’s Segmentation | 32 |
| Image type comparison for semantic segmentation | 32 |
| Pos-Processing on segmented images and Marker-Controlled Watershed Segmentation | 33 |
| Classifier-based filtering and false positive reduction | 34 |
| Feature extraction | 36 |
| Conclusions | 37 |
| References | 61 |
| CHAPTER THREE: COMPUTER VISION SYSTEM FOR BEEF CATTLE WEIGHT PREDICTION IN COMMERCIAL FEEDLOTS: INNOVATIONS IN IMAGE FEATURE EXTRACTION..... | 65 |
| Abstract | 65 |
| Introduction | 66 |
| Material and Methods | 68 |

| | |
|--|-----|
| Dataset..... | 68 |
| Data analysis | 69 |
| Fourier Transform | 70 |
| Ridge and Lasso Regression | 72 |
| Partial Least Squares | 73 |
| Artificial Neural Network (ANN) | 74 |
| Results and Discussion | 74 |
| Conclusions | 79 |
| References | 88 |
| | |
| CHAPTER FOUR: EXPLORING DEEP LEARNING FOR FACE-BASED IDENTIFICATION IN BEEF CATTLE CONSIDERING TWO DIFFERENT LOCATIONS OF THE PRODUCTION SYSTEM..... | 91 |
| Abstract | 91 |
| Introduction | 92 |
| Material and Methods | 94 |
| Data Acquisition..... | 94 |
| Data Analysis | 96 |
| Object Detection and YOLO | 97 |
| Animal Identification via Convolutional Neural Networks (CNN)..... | 98 |
| Results and Discussion | 100 |
| Conclusions | 103 |
| References | 116 |
| | |
| CHAPTER FIVE: CONCLUDING REMARKS..... | 119 |

LIST OF TABLES

| | | |
|-----------|--|----|
| Table 2.1 | Number of correct segmentation and its percentual | 39 |
| Table 2.2 | Performance of image classification using AlexNet and ResNet-50 with support vector machine (SVM) on different image types | 39 |
| Table 2.3 | Impact of data augmentation on AlexNet Image Classification using the combined image converted to jet colormap | 40 |
| Table 2.4 | Performance of AlexNet classifier trained with synthetic images on real and unseen images during training | 41 |
| Table 2.5 | Performance of semantic segmentation on different image types | 42 |
| Table 2.6 | Performance of ResNet50 in Classifying Images of Animals: Headless Body vs. Headed Body | 42 |
| Table 2.7 | Summary of Feature Statistics for Top-View Animal Images | 42 |
| Table 2.8 | Intersection over Union (IoU) measures between the predicted segmented images and the ground truth segmented images | 43 |
| Table 2.9 | Correlation between biological features of our segmented images with biological features of gold standard images | 43 |
| Table 3.1 | Descriptive statistics of body weight of beef cattle in feedlot | 81 |
| Table 3.2 | Performance of weight prediction models for beef cattle using different image features | 81 |

LIST OF FIGURES

| | | |
|-------------|--|----|
| Figure 2.1 | (A) Depth image equipment consisting of RFID antennas to trigger the 3D camera when an animal approaches to the water source. (B) An example of a top-down view image collected by our equipment..... | 44 |
| Figure 2.2 | (A) The Intel® RealSense™ D435 camera used in our experiment. (B) Types of images that the 3D camera provides..... | 45 |
| Figure 2.3 | Key questions for the challenges discussed in this article..... | 46 |
| Figure 2.4 | Composition of an infrared and depth images..... | 47 |
| Figure 2.5 | Flowchart depicting the selection process of the most efficient model for different combinations of neural networks and image types..... | 48 |
| Figure 2.6 | (A) AlexNet network architecture (adapted from Babu, 2016). (B) ResNet network architecture (adapted from Almabdy and Elrefaei, 2018)..... | 49 |
| Figure 2.7 | (A) Training of the GAN generator network. (B) Training of the GAN discriminator network..... | 50 |
| Figure 2.8 | (A) Steer binary image, and (B) Surface of the steer binary image. (C) Steer binary image, and (D) Distance transform of the steer binary image. (E) Distance transform of the steer image's complement, and (F) Negative of the distance transform of the steer image's complement..... | 51 |
| Figure 2.9 | (A) Steer over-segmentation of Watershed algorithm, (B) Steer Marker-Controlled Watershed Segmentation, (C) Body separated from the other parts after Marker-Controlled Watershed Segmentation, and (D) Final segmentation – body's steer..... | 52 |
| Figure 2.10 | (A) Biological features from the steer. The bounding box (yellow), contour/perimeter (green), major length (blue line), minor length (red line). (B) Measuring lines for width (red line) and length (blue line)..... | 53 |
| Figure 2.11 | Examples of depth and infrared image segmentation and their respective combinations..... | 54 |
| Figure 2.12 | Evaluation of GAN model performance. (A) usable synthetic images generated by the GAN model using the combined image. (B) Score of generated images by the generator (green) and discriminator (orange) over interactions. (C) usable synthetic images generated by the GAN model using colorjet image. (D) Score of generated images by the generator (green) and discriminator (orange) over interactions..... | 55 |

| | | |
|-------------|--|-----|
| Figure 2.13 | Sample of 25 synthetic images generated from the colorjet image, including four categories of unusable images. (A) Two animals. (B) Multiple animals. (C) Partial views of one animal. (D) background or image error..... | 56 |
| Figure 2.14 | The worst and the best semantic segmentation case for each type of image..... | 57 |
| Figure 2.15 | Top-view images showcasing feedlot cattle image diversity in terms of position, illumination, and lack of pixels in the images..... | 58 |
| Figure 2.16 | Examples of the output for semantic segmentation and our image processing technique for improving its image quality..... | 59 |
| Figure 2.17 | Examples of selected and discarded images in classifier-based filtering and false positive reduction..... | 60 |
| Figure 3.1 | Frequency distribution of images for each of the 79 animals in our experiment..... | 82 |
| Figure 3.2 | (A) Binary image of a square. (B) Magnitude spectrum of the Fourier Transform for the binary image of the square..... | 83 |
| Figure 3.3 | (A) Binary image of the animal's body from Top View. (B) Magnitude spectrum of the fourier transform for the binary image of the animal's body..... | 84 |
| Figure 3.4 | Top-view image of a beef cattle with 20 width and length measurements used as predictors in a multiple linear regression model with real weight as the response variable. Gray highlighting indicates the significant Pearson correlation between the feature and the actual weight ($P < 0.05$). (A) Width measurements; (B) Length measurements..... | 85 |
| Figure 3.5 | Correlation between observed and predicted body weight using Fourier features as predictors in the ANN model..... | 86 |
| Figure 3.6 | (A) Example of a magnitude spectrum profile for an animal body image. (B) Six theoretical ellipsoidal shapes, ranging from a 'flatter' ellipse to a more 'elongated' one, while keeping the size of the major axis constant. (C) Magnitude spectrum profiles for the six ellipsoidal shapes | 87 |
| Figure 4.1 | Video collection example at two distinct points in the industry: (A) Location 1, the exit of animals from the feedlot. (B) Location 2, the entrance of animals to the slaughterhouse | 105 |
| Figure 4.2 | Example of Yolo network applied to videos of Location 1 and Location 2. (A) Face detection (B) Cropped images in Location 1. (C) Cropped images in Location 2..... | 106 |

| | | |
|-------------|--|-----|
| Figure 4.3 | Overview of CNN architecture for beef cattle image identification in two locations of the livestock industry..... | 107 |
| Figure 4.4 | Examples of Nelore and Nelore crossbreed included in our models..... | 108 |
| Figure 4.5 | YOLOv8 performance in cattle face detection. (A) F1-confidence curve. (B) Precision-recall curve. (C) Loss and Mean Average Precision (mAP)..... | 109 |
| Figure 4.6 | Frequency distribution of detected image's face per animal..... | 110 |
| Figure 4.7 | Example of repeated images of faces of the same animal in the experiment. (A) Location 1. (B) Location 2..... | 111 |
| Figure 4.8 | Accuracy of RN50 and XPT for animal classification considering the same location (slaughterhouse) for both training and test image sets and considering images generated at a different location (feedlot) as testing set..... | 112 |
| Figure 4.9 | Heatmaps of confusion matrices for animal identification using location 2 (slaughterhouse) data as the training set. (A) RN50 with test set at location 1 (feedlot). (B) XPT with test set at location 1 (feedlot). (C) RN50 with Test Set at Location 2 (Slaughterhouse). (D) XPT with Test Set at Location 2 (Slaughterhouse)..... | 113 |
| Figure 4.10 | Examples of animals accurately identified at Location 1 when trained with images from Location 2 using the RN50 model..... | 114 |
| Figure 4.11 | Examples of animals accurately identified at Location 1 when trained with images from Location 2 using the XPT model..... | 115 |

CHAPTER ONE: INTRODUCTION

The beef industry plays a crucial role in supplying high-quality protein in response to the growing global demand. In 2023, the estimated global production of beef reached approximately 59.1 million tons, with significant increases in key nations, including the United States, Australia, Brazil, China, and India (OECD/FAO, 2023). Beyond its significance as a food source, the beef industry holds substantial economic impact, generating employment opportunities and providing support to rural communities. This economic interdependence extends across a network of stakeholders, ranging from producers to processors, distributors, and retailers, all playing a pivotal role in the global trade landscape (Smith et al., 2018).

In the ever-evolving landscape of the beef industry, strategic approaches have become essential for the sustainability of beef cattle production, for example the application of modern technologies to improve the efficiency of the final stage of beef production, i.e., feedlots. These specialized facilities play a pivotal role in the journey from pasture to plate. At the feedlots, animals enter their final phase of growth and fattening to get ready for market. In modern feedlots, the growth and development of animals should be closely monitored to guide management practices, such as daily feed delivery, detection of sick animals, and optimal harvest time.

Traditional weighing scales are commonly used to measure body weight (BW) in cattle, but periodic weighting is stressful for the animals, costly, and labor-intensive (Schofield et al., 1999; Cooke, 2014; Haskell et al., 2014; Lees et al., 2020). Consequently, cattle weight is typically assessed only at the beginning and end of the fattening phase, and management decisions during the finishing phase rely on subjective, visual appraisal from experts. An alternative in this regard refers to electronic scales strategically placed in each pen to passively measure weight, for example each time an animal visits the water tank. While these devices are very accurate, their cost and

maintenance requirement are generally prohibitive for commercial operations. In this context, a potential stress-free solution for BW measurement is the use of computer vision techniques based on animal images.

In the field of livestock, computer vision has emerged as a promising tool for predicting BW in production animals (Fernandes et al., 2019; Cominotte et al., 2020; Oliveira et al., 2021; Dorea and Rosa, 2022; Hou et al., 2023). These pioneering studies have demonstrated the considerable potential of weight prediction based on images, particularly in controlled and confined environments. However, a significant challenge lies in extending these findings to feedlot applications, where animals roam freely within pens, and in outdoors condition in which various environmental factors can influence the accuracy of weight predictions.

Beyond monitoring animal weight in feedlots, the use of cattle identification throughout the production process is essential for farmers and producers. This process, commonly referred to as 'animal identification' or 'ID,' involves maintaining detailed records of individual farm animals or groups to facilitate tracking from birth to marketing (Greene, 2010). It serves a variety of crucial purposes, including internal management tasks like maintaining health records, tracking vaccination history, recording genetic information, and age determination, as well as enabling the tracking of animal movements between different premises.

A thorough comprehension of the origin, movement, and individual identification of animals within the production chain is indispensable for guaranteeing product quality and safety (Velthuis et al., 2003), preventing disease outbreaks through rapid and precise identification of affected animals (Smith and Sanderson, 2023), optimizing logistics efficiency in cattle transportation within the production chain (Ljungberg et al., 2006; Miranda et al., 2014), and ensuring strict compliance with industry regulations (Bruneau et al., 2021).

While the industry has made significant investments in modernizing animal identification through Technology in Traceability Management (TTM) programs using RFID devices, these technologies can be costly and pose challenges in terms of maintenance and management. Additionally, they are susceptible to being lost, removed, or damaged (Gaber et al., 2016). Another drawback of this method is that once an animal is slaughtered, morphological traits are lost, and the RFID tag is separated from the carcass. This can lead to information loss, interruptions, or confusion in the subsequent stages of the supply chain, compromising the reliability of traceability (Zhao et al., 2019).

In this context, computer vision (CVS) emerges as an alternative for animal identification, with diverse applications primarily in beef cattle (Bezen et al., 2020; Kumar et al., 2018; Qiao et al., 2019, Li et al. 2022) and swine industries (Hansen et al., 2018; Marsot et al., 2020; Zheng et al., 2015). These approaches aim to leverage unique individual characteristics, such as body patterns, facial features, muzzles, or iris patterns. Notably, none of these studies have specifically addressed the problem of identifying the same animal through facial images captured at various locations along its journey within the meat production chain.

Hence, this dissertation explores the application of computer vision to address challenges in beef cattle production, which include: 1) Developing a comprehensive image processing pipeline to collect, process, and prepare top-view images of beef cattle in outdoor feedlot environments for future animal weight predictions. This involves image preprocessing, image classification to identify images with a single animal, and techniques for extracting the animal's body from the frame. 2) Developing an additional module within the previously mentioned pipeline to predict body weight (BW) by utilizing features extracted from the body images of the animals. In this regard, our objective was twofold. Firstly, we conducted a comprehensive and

comparative evaluation to assess the effectiveness of Fourier feature extraction techniques in comparison to traditional image-derived features. Secondly, we explored and compared the performance of four machine learning techniques (partial least squares regression, ridge regression, lasso regression, and artificial neural networks) for predicting the body weight (BW) of beef cattle in commercial feedlot environments. 3) Investigating methods for animal identification conducting a study to explore the potential of traditional neural networks in the identification of animals within and across two distinct locations in the livestock industry using its facial images.

References

- Bezen, R., Y. Edan, I. Halachmi. 2020. Computer vision system for measuring individual cow feed intake using RGB-D camera and deep learning algorithms. *Comput. Electron. Agric.* 172:105345. doi:10.1016/j.compag.2020.105345.
- Bruneau, J., and A.I Ugochukwu. 2021. Consumer welfare of country-of-origin labelling and traceability policies. *Agronomy*.11:916. doi: 10.3390/agronomy11050916.
- Cominotte, A., A. F. A. Fernandes, J. R. R. Dorea, G. J. M. Rosa, M. M. Ladeira, E. H.C.B. Van Cleef, G. L. Pereira, W. A. Baldassini, and O. R. Machado Neto. 2020. Automated computer vision system to predict body weight and average daily gain in beef cattle during growing and finishing phases. *Livest Sci.* 232:103904–14. doi: 10.1016/j.livsci.2019.103904.
- Cooke, R. F., and E. Bill. 2014. Kunkle Interdisciplinary Beef Symposium: Temperament and acclimation to human handling influence growth, health, and reproductive responses in *Bos taurus* and *Bos indicus* cattle. *J. Anim. Sci.* 92:5325–5333. doi:10.2527/jas.2014-8017.
- Dac H. H., C. Gonzalez Viejo, N. Lipovetzky, E. Tongson, F. R. Dunshea, and S. Fuentes. 2022. Livestock identification using deep learning for traceability. *Sensors* 22:8256. doi: 10.3390/s22218256.
- Dorea, J. R. R., and G. J. M. Rosa. 2022. Computer vision systems to advance high-throughput phenotyping in livestock. *Proc. 12th World Congress on Genetics Applied to Livestock Production (WCGALP)*. 138: 602 – 605.
- Fernandes, A. F. A., Dórea, J. R. R., Fitzgerald, R., Herring, W., and Rosa, G. J. M. 2019. A novel automated system to acquire biometric and morphological measurements and predict body weight of pigs via 3D computer vision. *J. Anim. Sci.* 97:496–508. doi:10.1093/jas/sky418.
- Gaber, T., A. Tharwat, A. E. Hassanien, and V. Snasel. 2016. Biometric cattle identification approach based on Weber’s Local Descriptor and AdaBoost classifier. *Comput. Electron. Agric.* 122:55–66. doi: 10.1016/j.compag.2015.12.022.
- Golan, E., B. Krissoff, F. Kuchler, L. Calvin, K. Nelson, and G. Price. 2004. Traceability in the US food supply: economic theory and industry studies. *Agricultural Economic Report*. 830:1-46.
- Greene, J. L. 2010. Animal Identification and Traceability: Overview and Issues. Report, November 29, 2010; Washington D.C. University of North Texas Libraries, UNT Digital Library. Accessed on November 9, 2023. URL: <https://digital.library.unt.edu/ark:/67531/metadc491014/>.

- Hansen, M. F., M. L. Smith, L. N. Smith, M.G. Salter, E. M. Baxter, M. Farish, and B. Grieve. 2018. Towards on-farm pig face recognition using convolutional neural networks. *Comput. Ind.* 98:145–152. doi:10.1016/j.compind.2018.02.016.
- Haskell, M. J., G. Simm, and S. P. Turner. 2014. Genetic selection for temperament traits in dairy and beef cattle. *Front. Genet.* 5:368. doi:10.3389/fgene.2014.00368.
- Hou, Z., L. Huang, Q. Zhang, and Y. Miao. 2023. Body weight estimation of beef cattle with 3D deep learning model: PointNet++. *Comput. Electron. Agric.* 213:108184. doi:10.1016/j.compag.2023.108184.
- Kumar, S., A. Pandey, K. Sai Ram Satwik, S. Kumar, S. K. Singh, A. K. Singh, and A. Mohan. 2018. Deep learning framework for recognition of cattle using muzzle point image pattern. *Measurement* 116:1–17. doi:10.1016/j.measurement.2017.10.064.
- Lees, A. M., H. E. Salvin, I. G. Colditz, and C. Lee. 2020. The influence of temperament on body temperature response to handling in Angus cattle. *Animals* 10:172. doi:10.3390/ani10010172.
- Li, G., G. E. Erickson, and Y. Xiong. 2022. Individual beef cattle identification using muzzle images and deep learning techniques. *Animals* 12:1453. doi:10.3390/ani12111453.
- Ljungberg D., G. Gebresenbet, and S. Aradom. 2006. Logistics chain of animal transport and abattoir operations. *Biosyst. Eng.* 96:267-277. doi: 10.1016/j.biosystemseng.2006.11.003.
- Marsot, M., J. Mei, X. Shan, L. Ye, P. Feng, X. Yan, C. Li, and Y. Zhao. 2020. An adaptive pig face recognition approach using convolutional neural networks. *Comput. Electron. Agric.* 173: 105386. doi: 10.1016/j.compag.2020.105386.
- Miranda-de la Lama G.C., M. Villarroel, and G. A María. 2014. Livestock transport from the perspective of the pre-slaughter logistic chain: a review. *Meat Sci.* 98:9-20. doi:10.1016/j.meatsci.2014.04.005.
- OECD/FAO. 2023. OECD-FAO Agricultural Outlook 2023-203. OECD Publishing. doi:10.1787/08801ab7-em.
- Oliveira, D. A. B., L. G. R. Pereira, T. Bresolin, R. E. P. Ferreira, and J. R. R. Dorea. 2021. A review of deep learning algorithms for computer vision systems in livestock. *Livest. Sci.* 253:104700. doi: 10.1016/j.livsci.2021.104700.
- Qiao, Y., D. Su, H. Kong, S. Sukkarieh, S. Lomax, and C. Clark. 2019. Individual cattle identification using a deep learning based framework. *IFAC-PapersOnLine.* 52:318–323. doi:10.1016/j.ifacol.2019.12.558.

- Schofield, C. P., J. A. Marchant, R. P. White, N. Brandl, and M. Wilson. 1999. Monitoring pig growth using a prototype imaging system. *J. Agric. Eng. Res.* 72:205–210. doi:10.1006/jaer.1998.0365.
- Smith M.R., and M.W Sanderson. 2023. Modeled impacts of rapid and accurate cattle tracing in a Foot-and-Mouth Disease outbreak in the US. *Prev. Vet. Med.* 215:105911. doi: 10.1016/j.prevetmed.2023.105911.
- Smith, S. B., G. Takafumi, and P. L. Greenwood. 2018. Current situation and future prospects for global beef production: overview of special issue. *Asian-Australas. J. Anim. Sci.* 31:927–932. doi: 10.5713/ajas.18.0405.
- Velthuis A. G. J., L. J. Unnevehr, H. Hogeveen, and R. B. M. Huirne. 2003. *New approaches to food safety economics*. 1st ed. 1 Springer Dordrecht.
- Zhao J., A. Li, X. Jin, and L. Pan. 2020. Technologies in individual animal identification and meat products traceability. *Biotechnol. Biotechnol. Equip.* 34(1): 48-57, doi: 10.1080/13102818.2019.1711185
- Zheng, S., S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. S.Torr. 2015. Conditional random fields as recurrent neural networks. *Proc. International Conference on Computer Vision (ICCV)*.1:1529–1537. doi:10.48550/arXiv.1502.03240.

CHAPTER TWO: A PIPELINE APPROACH FOR 3D IMAGE PRE-PROCESSING,
CLASSIFICATION, SEGMENTATION, AND FEATURE EXTRACTION FOR FUTURE
WEIGHT PREDICTION OF BEEF CATTLE IN FEEDLOTS

Abstract

This paper explored the feasibility of using computer vision technology in feedlot conditions by capturing top-down 3D images of 150 animals during their water intake sessions over three months. However, implementing this technology outdoors with unrestrained animals presents challenges, including harsh equipment conditions, variable lighting, and animal movement. In response, we proposed a comprehensive pre-processing framework consisting of four key modules. First, we aimed to enhance image quality and emphasize the animal within the scene. Second, our focus was on identifying the optimal combination of image types (depth, colorized depth, infrared, and infrared-depth combination) and neural network architectures for accurate classification of images containing a single entire animal (usable images). Third, we evaluated the effectiveness of data augmentation using synthetic images to reduce false positives in classification. Lastly, we investigated segmentation techniques to extract only the animal body from the image (excluding head and tail). Our results demonstrated the effectiveness of combining infrared and depth images for noise reduction, achieving correlations in the range of 0.60-0.90 with gold standard segmented images. The utilization of the AlexNet classifier with these combined images achieved an accuracy of 0.966. Incorporating 500 augmented synthetic images generated by GANs significantly enhanced classifier performance, reducing the false positive rate from 0.031 to 0.008. Applying the Marker-Controlled Watershed Segmentation (MCWS)

effectively isolates the animal body in images, achieving an average overlap of 88.8% with ground truth segmented images. The integration of these modules paved the way for future automated computer vision systems in feedlots.

Introduction

The relatively small profit margins and high production costs incur in increased risk of economic losses in beef cattle feedlots. In this context, optimized management decisions are central key for feedlot profitability, for example, efficient feed management and harvest time decision. These two components of management are directly related to the body weight and daily gain of cattle. However, feedlot cattle are usually weighed only at the beginning and end of the finishing phase. As such, animals with unsatisfactory growth performance, e.g., animals that get sick, may not be detected on a timely fashion. On the other hand, faster growth animals will finish earlier but unnecessary extended time on feed will reduce profitability. Thus, the use of tools to access body weight and conformation in real-time can be a powerful alternative to improve operational decisions by early detecting inefficient animals, improving feed management, and determining optimal harvesting time.

From a logistic point of view, weighing animals in a frequent basis is unfeasible due to animal stress and labor requirement. A strategy to avoid these two issues is the use of electronic scales in each pen for passively weighing animals, for example when they visit the water source. Such scales however are relatively expensive, and require routine maintenance. More recently, an alternative that has been successfully explored is the use of computer vision systems (CVS) to extract animal biometric measurements from three-dimensional (3D) images (e.g., Gomes et al., 2016; Kashiha et al., 2014; and Wang et al., 2008), from which body weight (BW) and

conformation can be assessed (e.g., Gomes et al., 2016; Li et al., 2014; and Ozkaya et al., 2016). Another recent article was published by Cominotte et al. (2020), who employed an automated CVS to predict body weight (BW) and average daily gain in beef cattle during growing and finishing phases.

The above-mentioned approaches, however, were developed under fairly controlled conditions, including restrained animals, indoor location, absence of natural light interference, and relatively small data sets, both in terms of number animals and timepoints. In this paper, we conducted a pilot study to assess the feasibility of estimating the weight of animals in real feedlot conditions. Using a 3D camera, we captured images of 150 animals over a three-month period in feedlot, subjecting our study to various challenges and natural environmental factors. Even though this 3D camera can be used outdoor, several factors can affect infrared light, thus reducing the quality of the images, including the angle of incidence, distance from the material, ambient light, among other (Zhong et al., 2019). In addition, depth images can experience a massive loss of depth information due to glare and reflection, especially on objects which have a high level of transparency, brightness, opacity, and absorbency (Alhwarin et al., 2014). Unfortunately, such complications are inevitable when images are collected in commercial production environments, as in this study, with variations in temperature and light throughout the day, presence of shadows, complex background, changes in animal posture, multiple animals in the same image, etc. All these issues make unfeasible or extremely difficult to apply standard digital image processing to such images.

The challenges and limitations of using CVS in commercial beef cattle feedlot environments underscore the importance of innovative and effective approaches to image processing and analysis. In this context, it is noteworthy that certain aspects remain relatively

uncharted territory. One such underexplored area pertains to the utilization of various image outputs generated by the Intel® RealSense™ Depth Camera D435, including depth, colorized depth, and infrared imagery. While color images are heavily used in Computer Vision and depth images have shown to be robust in many with controlled light to the best of our knowledge this is the first study to attempt to optimize the use of multiple sources of images from depth cameras. This study seeks to address this gap by evaluating the performance of two prominent image classifiers, AlexNet and ResNet-50 in conjunction with Support Vector Machine (SVM) across distinct image types, including depth, colorized depth, infrared, and composed (infrared-depth combination).

We also encountered the challenge of having a limited number of available images for classifier training. To address this, we took an innovative approach, inspired by the successful use of Generative Adversarial Networks (GANs) in image classification, as demonstrated in recent research (Zak, Jakub, et al., 2022; Frid-Adar et al., 2018; Lu et al., 2022; Bird et al., 2022). By employing GANs to create synthetic images, our goal is to improve the effectiveness and accuracy of our classifier. This aligns with the current trend of advancements in generative models, such as GPT-3, Google Bard, and other Large Language Models (LLMs). Ultimately, our aim is to create a more diverse and robust dataset without the need for additional real-world experiments.

Additionally, a noteworthy question we addressed in this article is the need to automatically isolate only the body of the animal from the images, excluding the head and tail. This refinement is particularly crucial in the context of future weight prediction, where potential biases due to variations in animal positioning during image capture must be mitigated. To achieve this, we employed Marker-Controlled Watershed Segmentation, a technique that focuses exclusively on the body segmentation process.

Therefore, the main objective of this paper was to establish a foundational framework for the future development of an image processing system aimed at automating weight prediction in feedlots. This system encompassed advanced techniques for enhancing image quality, accurately classifying scenes with only one single animal, and segmenting only the animal's body in the image (without head and tail). With this, we aimed to achieve the overarching goal of developing an efficient pipeline with four sub-objectives as defined below.

The first sub-objective was to perform image preprocessing to address issues related to lighting heterogeneity, enhance the visibility of the animal body while reducing background interference, and rectify distortions such as pixel gaps or 'holes' in the images. The second sub-objective was to identify the optimal combination of image type and neural network architecture for accurately classifying images as 'usable' (those containing only one whole animal in the scene) or 'unusable' (only parts of one animal, multiple animals, or background). The third sub-objective was to determine if the use of data augmentation with synthetic images can effectively improve the performance of classification, aiming to reduce the false positive rate (images classified as 'usable' but were unusable). The fourth sub-objective was to investigate potential differences in segmentation quality among the available image types using semantic segmentation to extract only the animal's body (without head, tail, and background) from the images selected by the best classifier.

Material and Methods

Data acquisition

The data set was collected at DSM Innovation and Applied Center, in Brazil. A total of five 3D cameras (Intel® RealSense™ Depth Camera D435) were used. The Intel® RealSense™ D435 camera is a USB-powered depth camera that features two infrared sensors (used to construct the depth image), an RGB sensor, and an infrared projector (Figure 2.2A). These features allow the camera to capture both color and depth images, as well as infrared images (Figure 2.2B). This type of camera has several advantages over conventional cameras. They can provide depth information, which allows for the creation of precise 3D models of objects and scenes.

The cameras were installed on the top of the water tank of five pens and were coupled with radio-frequency identification (RFID) readers to detect the presence of an animal in the camera field. Whenever an animal came to the water tank, the Ultra High Frequency (UHF) RFID antenna read its electronic ear tag and simultaneously triggered the camera to capture providing top-down view images of the animal (Figure 2.1). Hence, each time an animal came to the water tank, its unique animal identification number (electronic ear tag) and a series of image frames were all uploaded wirelessly to the cloud database. The cameras captured images of each animal for approximately 90 days generating a dataset comprising 17,754 images of 150 cattle in total. Additionally, electronic scales were used to measure and record the actual weight of the animals.

Image processing and Computer Vision pipeline

The main objective of this paper was to establish a foundational framework for the future development of a computer vision system aimed at automating weight prediction in feedlots. This

framework included advanced techniques for enhancing image quality and precise image classification, identifying scenes with only a full-body animal, as well as segmenting the animal body while excluding the head, tail, and background. To achieve this, we proposed four sub-objectives, as described below:

1. **Image Pre-Processing:** perform image preprocessing to address issues related to lighting heterogeneity, enhance the visibility of the animal body while reducing background interference, and rectify distortions such as pixel gaps or 'holes' in the images.
2. **Image classification:** identify the optimal combination of image type and neural network architecture for accurately classifying images as 'usable' (those containing only one complete animal in the scene) or 'unusable' (only parts of one animal, multiple animals, or background).
3. **Image Augmentation with Synthetic Images:** determine if the use of data augmentation with synthetic images can effectively improve the performance of classification, aiming reduce the false positive rate (classify usable as an unusable image) ensuring that the model is effective with real images.
4. **Animal Body's Segmentation:** investigate potential differences in segmentation quality among the available image types using semantic segmentation to extract only the animal's body (without head, tail, and background) from the images selected by the best classifier.

Figure 2.3 presents the key questions that motivated the development of each of the four steps (modules) in the methodology. Further details are provided in the following sections.

First Module: Image Pre-Processing

The infrared projector in the 3D camera is particularly useful in low-light conditions, as it is capable of capturing depth information even when visible light is limited. The depth camera is especially useful for applications where depth accuracy is important, such as in movement tracking, which is crucial for our experiment. Finally, the RGB camera is a conventional camera that captures high-resolution color images.

Even though this 3D camera can be used outdoor, several factors can affect infrared light, thus reducing the quality of the images, including the angle of incidence, distance from the material, ambient light, among others (Zhong et al., 2019). In addition, depth images can experience a massive loss of depth information due to glare and reflection, especially on objects which have a high level of transparency, brightness, opacity, and absorbency (Alhwarin et al., 2014). Unfortunately, such complications are inevitable when images are collected in commercial production environments, as in this study, with variations in temperature and light throughout the day, presence of shadows, complex background, changes in animal posture, multiple animals in the same image, etc. Therefore, aiming to overcome some of the limitations of individual image types, we decided to make a composition between depth and infrared images through the *imfuse* function from MATLAB (Release 2020b). The idea is to create a blended overlay image, by scaling the intensities of depth and infrared images jointly as a single data set and keeping an associated matte for each element and its pixel values. The more the two images share in terms of pixel intensity values, the darker will be the blended area. On the other hand, the less the images share on pixel intensity values, the greener will be that specific area. In the end, shadows, lost pixels, and most of the background will be underemphasized from the image, and the target object in the

scene (the steer in our case) will be highlighted in black (Figure 2.4). To verify the effectiveness of this approach, we conducted a detailed visual inspection by binarizing 60 depth images, 60 infrared images, and 60 composite images to assess the completeness of object segmentation in each image type.

Second Module: Image classification

The second step in the process was to classify each image as usable or unusable. The input dataset included 17,754 images for each type of image as described above and illustrated in Figure's 2 and 3. Usable images were defined as those containing one single animal captured within the frame with its entire body on the frame. Unusable images, on the other hand, were defined as those with only background (i.e., no animal at the frame), or only part of an animal body, or multiple animals (with partial or entire body) on the frame.

To classify images as usable or unusable, we utilized two popular image classification networks, namely AlexNet and ResNet-50, with support vector machine (SVM). It is important to note that both networks require a 3-channel image input. Therefore, in cases where a single-channel input image was used, it was necessary to duplicate the same image in all three channels to create a compatible input for the architecture. For depth images, we utilized functions like 'im2double' to automatically manage rescaling and offsetting, ensuring safeguarding the crucial depth information. Additionally, we applied traditional data augmentation techniques to enhance the robustness and performance of both classification models. The technique involved applying random transformations to the original images in the training dataset using the 'augmentedImageDatastore' function in MATLAB. These transformations included resizing, rotation, reflection, and color augmentation to introduce variations in the images. The goal was to

prevent the models from overfitting to the specific details of the training images, thereby promoting better generalization to new data. The size of the augmented dataset was increased approximately threefold compared to the original dataset. We selected the most efficient model from this combination of neural networks and image types based on metrics such as accuracy, precision, specificity, recall, F1-score, false positive, false negative, and running time, as described in Figure 2.5. Further details about the networks are described below.

AlexNet

AlexNet is a Convolutional Neural Network (CNN) consisting of three fully connected layers and five convolutional layers. The initial layer processes filtered images with dimensions of $227 \times 227 \times 3$. The architecture originally includes 1000 connected layers, primarily used for feature extraction, producing a 4096-dimensional feature vector for each input image. It comprises 650,000 neurons and 60 million parameters. In this study, we employed a modified version of AlexNet, pre-trained on the ImageNet dataset (Krizhevsky et al., 2017). We replaced its final layers with a fully connected layer for two classes, a softmax layer, and a classification layer. Training was conducted using the Stochastic Gradient Descent with Momentum (SGDM) optimizer over 7 epochs with a batch size of 32 to prevent overfitting. The initial learning rate was set at 0.001, and validation occurred every 5 epochs. We employed techniques like weight freezing, dropout, and data augmentation to enhance model generalization and mitigate overfitting (Gupta et al., 2021; Thamaraimanalan et al., 2019). Figure 2.6A illustrates the AlexNet architecture.

ResNet-50 with support vector machine (SVM)

ResNet-50 is a popular deep neural network for image classification due to its effectiveness. It comprises 50 layers and excels in image-related tasks. The key innovation in the ResNet architecture is the use of "skip connections" or "shortcut connections," which facilitate the smoother flow of information across layers, mitigating the vanishing gradient problem. In our approach, we utilized a pretrained ResNet-50 network as the foundation for our image analysis. To achieve precise image categorization, we integrated ResNet-50 with a Support Vector Machine (SVM). We extracted essential features from the 'fc1000' layer, positioned just before the classification layer, and utilized them as inputs for a linear SVM, a well-established supervised learning algorithm. The SVM effectively distinguishes between different image classes, resulting in substantial enhancements in classification accuracy and computational efficiency (Almabdy & Elrefaei, 2019). Figure 2.6B illustrates the architecture of ResNet-50. The ResNet-50 architecture is shown in Figure 2.6B.

Third Module: Image Augmentation with Synthetic Images

Generative Adversarial Networks

Generative Adversarial Networks (GANs) are a type of deep generative model, which falls under the category of generative neural network models. Generative models are designed to capture the underlying essence of data, encompassing statistical nuances and inherent structures. Within the image dataset comprising aerial views of cattle bodies, the underlying distribution might encapsulate diverse attributes like body shapes and sizes, coat colors, and even the distinctive positions of animals while drinking water. Deep generative models use deep neural networks with multiple layers to generate new data samples that are similar to a given training

dataset. Due to their ability to capture more complex patterns and structures in the data, deep generative models can generate more realistic new data samples (in our case synthetic images).

Generative Adversarial Networks (GANs) were developed by Goodfellow et al. in 2014 and find applications in various aspects of computer vision within the field of animal science, including data augmentation, image synthesis, image inpainting, and image super-resolution. In this study, we specifically utilized GANs for data augmentation (Zak, Jakub, et al., 2022; Frid-Adar et al., 2018; Lu et al., 2022; Bird et al., 2022), generating synthetic top-down view images of a cattle farm. This approach aims to create a more robust and diverse dataset of images for training classifiers.

GANs consist of two neural networks: a generator that creates images and a discriminator that evaluates them. They engage in adversarial training, with the generator attempting to produce images that are indistinguishable from real ones, while the discriminator's task is to distinguish between real and generated images.

This approach can be illustrated as a contest between a cop and a thief. Suppose the goal is to create a synthetic image of a top-down view of cattle. The generator, or 'synthetic image generator,' functions like a thief, generating an image with random noise and trying to trick the discriminator, or 'real/synthetic image detector.' The discriminator's goal is to classify each image provided by the synthetic image generator by determining whether it's real or synthetic and providing feedback on the likelihood of improvement.

Over time, the generator becomes skilled at creating synthetic images similar to real ones, and the discriminator can no longer distinguish between real and synthetic images. Figure 2.7 illustrates how the generator and discriminator networks are trained together in a GANs.

Generating synthetic images using GANs

To create synthetic images, we followed a multi-step process designed to enhance the realism and accuracy of the generated images. We initiated with a training dataset comprising both usable and unusable original images. For the usable images, we employed Generative Adversarial Networks (GANs) to generate synthetic counterparts, while for the unusable images, we adopted a more specialized approach. The unusable images were further categorized into four sub-labels ('two animals,' 'multiple animals,' 'partial views of one animal,' 'background or image error'). For each of these sub-labels, separate GANs were trained to refine the generation of synthetic images.

Our initial approach was to use GANs to generate synthetic images that resembled the type of images that performed best in our original classifier, which were the combined images. Additionally, we also attempted to mimic other types of images, such as only the infrared, only the depth, and composite images by inserting them into each of the three color channels. For example, we placed the infrared in the first channel, depth in the second channel, and the combined image in the third channel. We explored all possible combinations of these channels, including applying the jet colormap to these channels with the aim of highlighting important features within the animal images.

After running the GANs to generate imitations in various scenarios, we selected the type of image that was best imitated. From this group, we proceeded to cull poorly imitated images using a clustering approach. This step involved evaluating the quality of the imitations and eliminating images that were not well replicated. For instance, we removed images that lacked clear contours or shapes in the animal representation.

From the pool of the best-replicated images, we gradually introduced them into our classifier. We began by augmenting the dataset with 100 images in the initial three rounds and later

added 200 images in the final four rounds. Following an evaluation of the classifier's performance after each round, we chose the model that demonstrated the most significant reduction in false positives. Below, we provide an overview of these steps adopted:

1. Generate synthetic images emulating various types, including infrared, depth, composed (infrared and depth), as well as combinations of images in the three color channels, and application of the jet colormap.
2. Select the best imitated image type visually and discard poorly imitated ones using clustering and quality assessment.
3. Introduce the best-replicated images gradually into our classifier, progressively expanding the dataset.
4. Choose the model that achieved the most significant reduction in false positives through performance evaluation of the classifier.

Classifier Selection

The goal of this step was to select the best classifier for the usable and unusable images. To achieve this, we evaluated the performance of the classification model trained with the original dataset versus the dataset augmented with synthetic images.

We utilized multiple evaluation metrics to ensure a comprehensive analysis of the classifier performance. While accuracy is a common metric used to evaluate the performance of an image classifier, it can be misleading and fail to account for the distribution of classification errors. Therefore, we also analyzed other metrics such as precision, recall, F1 score, as well as the false positive and negative rates of the classifier.

By evaluating the overall performance of the classifier and considering multiple metrics, we were able to identify its weaknesses and guide our efforts towards improving its performance, thereby avoiding any misleading or inaccurate results. This approach allowed for the selection of the best classifier, ensuring its ability to generalize to new real images.

Fourth Module: Animal Body's Segmentation

Based on images containing only one animal obtained through our classification model, this section aimed to extract the animal from the scene while excluding the head and tail. Hence, for images classified as usable, we performed image processing in five steps. 1. Image type comparison for semantic segmentation (determine the best type of image to locate the pixels related to the animal body in the image); 2. Pos-processing on segmented images (improve the quality of the segmented images obtained through semantic segmentation); 3. Marker-Controlled Watershed Segmentation (remove the pixels from the animals' head and tail); 4. Classifier-based filtering and false positive reduction (discard images where the animal's head was not successfully removed and any other errors that may have been caused by possible classifier inaccuracies or failures in semantic segmentation); and 5. Feature extraction (extract biological features of the animal body for future BW prediction).

Image type comparison for semantic segmentation

In our study, we explored the performance of various image types, including depth, colorized depth, infrared, and the combination of infrared and depth, for semantic segmentation using a convolutional neural network (CNN) architecture inspired by the U-Net framework. This

architecture incorporates downsampling layers for feature extraction and corresponding upsampling layers for output reconstruction.

To train the network, we labeled 457 animal body images using the ImageLabeler tool in Matlab to distinguish between the animal body without head and the background. For each image type, 366 images were used for training, and 91 images were set aside for testing. The prediction accuracy was evaluated using a range of metrics, including global accuracy, mean accuracy, Intersection over Union (IoU), mean BFScore, and the Boundary F-1 score, along with a confusion matrix. These metrics were chosen as they provide insight into the accuracy and robustness of the segmentation model. These metrics will be further explained in the following paragraph.

Global Accuracy measures the overall accuracy of the segmentation model, which is calculated as the number of correctly classified pixels divided by the total number of pixels in the image. Mean Accuracy computes the average of the per-class accuracies, where each class is weighted equally. Intersection over Union (IoU) measures the overlap between the predicted segmentation and the ground truth segmentation, and it is calculated as the ratio of the intersection between the predicted and ground truth segmentations to their union. Mean BFScore is a boundary-based segmentation evaluation metric, which takes into account the boundary distance between the predicted and ground truth segmentations. Boundary F-1 score is another boundary-based evaluation metric, which combines precision and recall of the boundaries in the segmentation.

Pos-processing on segmented images

After selecting the image to be used for segmenting the images, in this step we subjected the images to post-processing in which we used intensity normalization techniques to minimize light variations between images using Gaussian filter. We also performed edge enhancement using

popular filters such as the Sobel filter. Finally, we filled gaps in the animal body using the *imfill* function in Matlab. These techniques were applied to the segmented images to improve their quality and to ensure that the biological features of the animals were used as reliable predictors in future weight prediction models.

Marker-Controlled Watershed Segmentation (MCWS)

In this step we aimed to get more accurate body dimension measures, disconnecting the animal's main body from its head and tail. Additionally, we intend to disconnect the animal from remaining background in cases where the image composition technique could not solve it completely. Therefore, in this step we aimed to extract only the animal body, without head, tail and any remaining background (Figure 2.8A). An interesting approach to separate objects that are connected in the same image is called watershed transform. To understand this procedure, a good strategy is to think of a steer image as a landscape (Figure 2.8B), where the bright yellow and dark colors represent higher and lower areas, respectively. The idea is to identify the catchment basin and its watershed lines, and then split them up.

One way to identify catchment basins in image processing is by using the distance transform (Gunilla, 1986). The distance transform of a binary image is the distance from every pixel to the nearest nonzero-valued pixel. For example, in the steer binary image on Figure 2.8C, the distance transform produces an image like Figure 2.8D. However, Figure 2.8D is not useful because only one catchment basin spans the entire image. A solution is to compute the distance transform of the image's complement (Figure 2.8E). Hence, to have a catchment basin for each object, we turned the two bright areas into catchment basins by multiplying the distance transform by -1 (Figure 2.8F).

Next, applying the Watershed Algorithm on Figure 2.8F image leads to an over-segmentation (Figure 2.9A) due to the presence of many local minima; each local minimum, no matter how small, becomes a catchment basin. For addressing this problem, Meyer and Beucher (1990) proposed the MCWS. The main goal of this approach is to adapt the traditional watershed algorithm to remove small intensity minima. In our image example, we suppressed minima in the distance transform whose depth was less than seven pixels to perform the watershed segmentation (Figure 2.9B). Therefore, we modified the distance transform to have only minima at the desired locations, and then we repeated the watershed steps above and obtained the body separated from the other parts (Figure 2.9C). After disconnecting the three parts, we selected the largest area, i.e., the body's steer (Figure 2.9D).

Hence, the steps for performing the MCWS were: 1) Compute D , the negative distance transform of the binary image complement obtained in the first segmentation; 2) Find in D the positive integers corresponding to the locations of each catchment basin for each object, called the label matrix L ; 3) Find the zero-valued elements of L , which are located along the watershed lines, to separate the objects in the original image; 4) Create a mask to suppress minima in D for which the depth was less than a threshold (t); 5) Modify D in order to have minima at the desired locations, and then repeat the steps 2-3; and 6) Select the largest area.

Classifier-based filtering and false positive reduction

This section addresses a significant challenge encountered when the watershed algorithm couldn't accurately separate animal heads from their bodies in the images. To tackle this issue, we introduced an additional classifier, specifically a ResNet50. To train the classifier, we categorized the images into two classes: 442 successful cases (images containing only the animal's body) and

442 unsuccessful cases (animal images with heads included). For the images classified as unsuccessful, we implemented an iterative process of adjusting the local minima value until only the animal's body (without the head) was retained.

Following the classifier-based filtering classifier, our methodology includes a critical step aimed at refining the dataset and reducing potential false positives. In this stage, we evaluated 457 gold label images by measuring specific features of the animal bodies, such as area and eccentricity, within the binary image generated through semantic segmentation. To ensure dataset accuracy, we applied threshold values to these features, defining acceptable lower and upper limits for area ($30354.75 \geq \text{area} \leq 60440.75$ pixels) and eccentricity ($0.883 \geq \text{ecc} \leq 0.957$ pixels). Any connected component meeting the threshold criteria was considered a valid animal body, while those failing to meet these criteria were removed as potential false positives. The threshold values were calculated using a formula similar to the one used in boxplots, involving the first quartile minus 1.5 times the interquartile range. This integrated process enhances dataset reliability, ensuring its suitability for future body weight analysis and modeling.

Feature extraction

Using a set of images containing only the full body of an animal without the head, obtained at the end of our segmentation process, we calculated several basic biological features such as body area, major axis length, minor axis length, equivalent diameter, and perimeter. We performed these calculations using the 'regionprops' function in MATLAB (Release 2021b). These features are commonly used in animal biometrics and can provide useful information about an animal's size and shape (Figure 2.10A). In addition, we also calculated various measures of length and width

along the body of the animal to explore their potential as predictors for future weight prediction models (Figure 2.10B).

Results and Discussion

First Module: Image Pre-Processing

In Figure 2.4, we observed some gaps in the animal body in the depth image, while the infrared image did not exhibit any degradation. Conversely, the infrared image displayed shading and variations in the animal's coat color, which posed challenges for standard digital image segmentation. In both cases, the resulting binary image was unsuitable for predicting body weight. However, by combining the two images, we successfully addressed these issues in both examples (Figure 2.11). This strategy streamlined our algorithm, making it more efficient without the need for complex decision steps. Moreover, using a combination of depth and infrared images allowed us to work with a significantly larger image sample compared to using either image type alone. During the visual inspection and comparison with the likely animal body, as defined in the methodology of this article, we found that the combination of depth and infrared images enabled the extraction of approximately 20% more whole animals than using the depth image alone (Table 2.1). These results are in line with the findings of Alhwarin et al. (2014), who had previously achieved positive results with this approach.

Second Module: Image classification

The results in Table 2.2 show that the AlexNet network was not a good option for classifying depth images like ours, with an accuracy of only 0.5. Although we applied filters to enhance the edges and correct lighting heterogeneity, the lower pixel intensity in this type of image

may have made it difficult for this architecture to distinguish between different objects or textures. According to our analyses, for this type of image, the ResNet-50 network is a better option, with an accuracy of 92.0%, precision of 92.5%, specificity of 91.5%, and recall and F1-score of 91.4% and 91.0%, respectively.

For infrared images, the AlexNet showed, on average, performance similar to that of the ResNet-50. This can be explained by the fact that although these images do not have much color information, the significant pixel intensities may have contributed to identifying distinct features in the objects.

It can be noted that the AlexNet showed superior performance on colored images (colored depth and combined), performing better on all metrics in the table. This result corroborates with the findings above, as mentioned by Buslaev et al. (2020), indicating that AlexNet effectively handles colors by incorporating changes in RGB channel intensities. This capability allows the network to better learn distinct features and results in improved performance.

Among all the evaluated combinations, the choice of the AlexNet classifier that uses combined images is justified since it presented the lowest false positive rate and was able to keep false negatives at acceptable levels, while also presenting other performance metrics close to or above 90%. The high precision value is also a positive indicator, as it indicates a high proportion of examples classified as positive that are actually positive, which is desirable when seeking to minimize false positives. This model provided us with 2100 usable images.

Third Module: Image Augmentation with Synthetic Images

Our first attempt was to produce synthetic images based on the image set of the group that showed the best performance in Table 2.2 - the combined images. However, the results were not

encouraging. As can be seen in Figure 2.12A, the synthetic images produced by the GAN model do not look realistic. Additionally, monitoring the score of the images produced by the generator and discriminator also showed poor performance throughout the interactions (Figure 2.12B).

To improve the amount of information received by the network, we decided to add colors to the images. Therefore, we converted the combined image into a colorjet color map, which resulted in very realistic images for both the usable and non-usable image labels, as evidenced in Figure 2.12C and Figure 2.13 (A-D). An advantage of this strategy is that the jet color map paints the images according to the intensity of light along the animal's body, providing more information for the network to learn and better represent the reality of the experiment. The GAN network produced in total, 25,000 synthetic images. This includes 5,000 usable images and 5,000 images for each of the sub-labels, images with different positions of the animals and different intensities of light. We could observe that the color information we provided to the network helped it learn better about the patterns in our images, enabling it to imitate them more accurately.

Although the synthetic images generated by our GAN network were well-imitated, as shown in the score graph generated by the generator and discriminator, the generation was not 100% accurate. Like any statistical model, there are errors in this process, and it is essential to select the synthetic images consciously to add only the best ones to the classifier's training data. Therefore, to refine the selection of synthetic images, we also took into account the level of blurriness and random noise in the images, aiming to exclude those that may not accurately represent the real-world objects we are trying to classify. In other words, not all generated images will be suitable for use, and a conscious selection process is crucial to ensure the best results for the classifier's training.

To address the issues related to the synthetic images generated by our GAN network, we adopted a feature extraction approach to capture the objects' shape. We calculated parameters such as area, perimeter, roundness, and eccentricities from the synthetic images, and then used clustering techniques to group them into five clusters for each label and sub-label. However, we found that two of the clusters for each label and sub-label contained images with compromised shapes or noise only. Thus, we discarded these clusters, resulting in approximately 20,000 images in total. Subsequently, we sampled these images in small quantities for each label and sub-label, and incorporated them into our classification model.

It is important to note that previous studies have shown improvements in classifier performance by incorporating synthetic images (Alsabei et al., 2021; Srivastav et al., 2021; Bargshady et al., 2022; Chatterjee et al., 2022; Hazra et al., 2022). This positive trend is also reflected in our work, where the inclusion of synthetic images significantly enhanced the classification model's performance, improving accuracy, precision, recall, and F1 score.

The first two attempts to increase the number of synthetic images yielded good results by reducing false negatives. However, the most interesting scenario occurred when 500 synthetic images were added, as the model achieved a significant reduction in the false positive rate from 0.03 to 0.008. After this point, the performance of the model seemed to stabilize. Beyond this point, it appeared that the model's performance reached a plateau. It was at this juncture that we made the decision to select this particular model. The reduction from 0.03 to 0.008 in false positives represents a significant advancement for our classifier, particularly in feedlot applications. With the ability to select images with a sensitivity greater than 93%, this is a very promising and novel approach in this field accurately and precisely.

Significantly, it is pivotal to underscore the meticulous curation process applied to the incorporation of GAN-generated images into training the classifier. Rather than indiscriminately adding all generated images, our approach centers on the judicious selection of images that faithfully replicate real-world data. This deliberate selection aims to enhance the classifier's ability to discern and learn from a broader spectrum of visual variations not originally present in the source images. This strategy fosters a greater diversity within the training dataset, contributing to improved classifier performance and adaptability across a wider range of scenarios.

It is crucial to note that achieving good results on the test set with augmented images does not necessarily imply that the model will perform well on real, unseen data. Therefore, to ensure the reliability and effectiveness of our approach, we evaluated the model's performance on real data that was not used during training. The results presented in Table 2.4 show that the model performed well on both augmented and real-world data, indicating that our approach is generalizable and reproducible, two essential qualities for a reliable classifier. These outcomes instill great optimism regarding the forthcoming implementation of Computer Vision Systems (CVS) in feedlots. They highlight the potential to diminish exclusive reliance on empirical experiments for acquiring diverse image datasets. Instead, we are poised to perpetually enhance our classifier's efficacy by harnessing synthetic data, suggesting reduced dependence on real-world data collection moving forward. This represents a significant advantage, as it allows us to optimize and refine classification models, leading to more accurate and efficient results. Therefore, at this stage, we decided to use the AlexNet classifier with the color-depth image, using 500 augmented synthetic images. This provided us with 1670 usable images which were in the next segmentation module.

Fourth Module: Animal Body's Segmentation

Image type comparison for semantic segmentation

The performance of semantic segmentation on colored depth images showed the highest overall performance in terms of global accuracy, mean accuracy, mean IoU, weighted IoU, and mean BFScore (Table 2.5). This image type also had the lowest false positive rate, indicating that it is more accurate at distinguishing animals from the background. The depth image type had the lowest performance among the four image types evaluated, with the lowest global accuracy, mean accuracy, mean IoU, weighted IoU, and mean BFScore. The infrared image type had a high level of accuracy, with a global accuracy of 0.942, mean accuracy of 0.859, mean IoU of 0.786, and weighted IoU of 0.895. It also had a low false negative rate, indicating its ability to detect animals even in low-light conditions. However, it had a higher false positive rate compared to the colored depth image type. The combined image type performed well overall, with a global accuracy of 0.918 and a mean IoU of 0.716. However, it had a higher false positive rate compared to the colored depth image type, indicating that it may not be as accurate at distinguishing animals from the background.

We can also visualize the worst and best-case scenarios of segmentation in Figure 2.14. Even in the worst-case scenario, the semantic segmentation for colored depth still managed to extract the animal from the image, which could be further improved through processing techniques. Although our semantic segmentation did not fully isolate the head from the body, it was in the colored depth image where the most effective removal of a portion of the head was observed in comparison to the other image types.

The fact that the semantic segmentation technique was not able to remove the head entirely from any of the selected images was expected for us due to the significant heterogeneity of

illumination, animal position variation, and holes in the images (Figure 2.15). These factors may require a larger training population than what was available in this study. Furthermore, it is worth noting that increasing the training population alone often does not fully address the challenges of semantic segmentation. As proposed in prior studies (Zhao et al., 2018; Ye & Han, 2021; Asano et al., 2022; Kwasniewska et al., 2019), post-processing is frequently necessary to enhance the output of semantic segmentation. Hence, given these outcomes, we opted to focus our efforts on enhancing these results through advanced image processing techniques, which will be discussed in the upcoming section.

Pos-Processing on segmented images and Marker-Controlled Watershed Segmentation

The first column of Figure 2.16 shows two examples of semantic segmentation output. We can observe that there are missing parts in the animal body, which blend with the green color of the background, making it difficult to extract the complete animal shape. However, by using detection and enhancement techniques for the animal edges, we were able to fill in these parts without interference from the background, resulting in the complete reconstruction of the body. This process was essential for extracting the biological features of the images and also prevented us from having to discard some of them. The last column displays an example of the animal image with the head excluded. To attain this outcome using MCWS, the value of 7 for detecting local minima in the distance image demonstrated the most effective performance. This selection was not arbitrary and was determined through experimentation with various values. We conducted a simulation by altering the local minima value across a range from 1 to 10, followed by a visual assessment of the results. We observed that very low values generated many markers and segments in the animal, while very high values generated few markers and segments. Out of a total of 1672 images, the MCWS method, when configured to suppress minima in the distance transform whose

depth was less than 7 pixels, successfully separated the head from the body in most cases. However, in 273 images (17%) where there were no abrupt transitions between the body, neck, and head of the animal, configuring the method with a value of 4 pixels proved to be more suitable for correct separation.

Classifier-based filtering and false positive reduction

Table 2.6 demonstrates that the ResNet50 classifier effectively differentiated images featuring only the animal's body from those including the complete animal (with its head) with high accuracy. These results remained consistent across both the test and validation sets. In the test set, the classifier achieved an accuracy of 0.983, with minimal false positives and false negatives, resulting in high precision, specificity, and recall. In the validation set, accuracy was slightly lower at 0.955, with a modest increase in false positives and false negatives, along with slightly lower precision, specificity, and recall scores.

Out of the 1670 images subjected to the MCWS algorithm, which involved suppressing minima in the distance transform with depths less than 7 pixels to separate the head from the body, 1288 (77%) were classified as images where MCWS successfully separated the head from the body, while 382 (23%) were classified as images where MCWS couldn't separate the head from the body. This outcome confirms our previous visual inspections. Subsequently, these 382 images underwent an additional round of processing using the MCWS algorithm, this time with a local minima value reduced to 4 pixels. This adaptation allowed us to recover an additional 142 (37%) images containing animals without heads. Consequently, we obtained a total of 1430 (86%) images featuring animals without heads through this iterative process. Table 2.7 presents a snapshot of the statistics of features that can best represent the top view of our animals. Analyzing the coefficient

of variation, we can see that the measures that have the higher variability in terms of animal shapes are area and eccentricity. Therefore, we decided to use these two measures to calculate their respective lower and upper cutoff limits and identify possible outliers in the image set and discard them. The implementation of this module in our code allowed for the selection of 1430 images out of 1286 to extract features, with 144 images being discarded.

To visually demonstrate the effectiveness of this processing technique, Figure 2.17 displays eight randomly selected images from each group. The results indicate that our filter indeed removed images that could not be improved by our processing technique, or that are false positives from our classifier. Interestingly, our MWCS algorithm is unable to separate the head from the body when the animal is highly curved. This finding is actually advantageous, as including animals with curved positions in our sample could compromise the accuracy of weight prediction, which is the final goal to be accomplished with the images. Furthermore, a significant contribution of this study is the introduction of the watershed algorithm to the academic community in the field of animal science, traditionally recognized for its prevalence in medical image processing, such as searching for lesions in breast regions (Xu et al., 2017 and Shen et al., 2021), as well as in cell segmentation (Lux and Matula, 2020).

The Intersection over Union (IoU) values presented in Table 2.8 offer valuable insights into the segmentation algorithm's performance. With a mean IoU of 0.888, we observe a reasonably good overlap in segmentation between our segmented images and the ground truth. Even in the least favorable scenario, characterized by a minimum IoU value of 0.749, a reasonable alignment with the ground truth segmented images is maintained. Furthermore, the presence of a maximum IoU value of 0.941 signifies a high degree of overlap with the ground truth in specific instances, highlighting the overall quality of the segmentation process. These findings hold particular

significance as they enable us to curate a set of 1286 images featuring animal bodies (with heads excluded). This dataset instills confidence in the extraction of valuable biological characteristics for future analysis.

Feature extraction

Our database consisted of 125 different animals. Throughout our experiment, we encountered various challenges, including infrastructure and electrical issues, which resulted in the temporary loss of tracking for 25 animals in the designated area. These tasks also affected our ability to consistently capture images during the 90-day data collection period. Out of the 17,754 total images collected, 1,430 images featured a single entire animal within the frame. This percentage of such images can be attributed primarily to the high sensitivity of the antenna, which occasionally triggered the camera prematurely, resulting in partial body capture. Additionally, this sensitivity occasionally led to empty frames when animals approached the antenna closely, inadvertently activating the camera. Furthermore, our observations revealed that animals often approached the water source in pairs or groups, resulting in images with multiple animals in the frame.

Nevertheless, despite these challenges, our dataset remained robust. Among the 1430 images containing a single entire animal, 40.8% of the animals had a limited number of images (1-5 images), 33.6% had a moderate representation (6-15 images), and 25.6% possessed a substantial volume of images (more than 15 images). Remarkably, over 80% of these animals had at least one image captured per month, providing a significantly richer dataset compared to what is currently accessible in feedlots for weight measurements. Looking forward, we are confident that small refinements in our data collection system can effectively address the issues we have highlighted.

For instance, adjustments in antenna sensitivity, strategic camera placement, and optimizing camera angles are all strategies we are considering for future improvements.

Table 2.9 presents Pearson correlations between the biological characteristics of the images segmented by the method employed and the biological characteristics of the images considered the "gold standard". The results reveal that most of the evaluated biological characteristics exhibit high and positive correlations (above 0.9), indicating that these traits were effectively captured by the segmentation method. Specifically, area, centroid coordinates (X and Y), equivalent diameter, and minor axis exhibited the highest correlations. However, some characteristics displayed lower correlations (below 0.8), such as major axes, eccentricity, perimeter, and circularity, suggesting that the segmentation method may not have accurately captured these traits.

These statistical results highlight the effectiveness of our approach in capturing relevant biological characteristics in images, providing solid evidence of its accuracy and utility in animal monitoring applications.

Conclusions

In conclusion, our study represents a critical step towards the development of an advanced image processing system aimed at automating weight prediction in confined animals. The primary objective of this research was to establish a robust foundational framework that encompasses a suite of sophisticated techniques specifically crafted to address the unique challenges of this domain. While semantic segmentation played a crucial role in our approach, it became evident that in the context of our study, it alone was insufficient to achieve the desired level of precision and information extraction.

Through extensive experimentation with a meticulously curated dataset comprising 17,754 images of 150 cattle, continuously monitored throughout their three-month tenure in the feedlot, we have achieved substantial progress. Our framework excels in enhancing image quality, enabling precise classification, accurately identifying scenes featuring single complete animals, and seamlessly segmenting only the animal's body from the scene (without head and tail).

These accomplishments underscore the potential for revolutionizing animal monitoring in confined environments, with applications extending beyond weight prediction. By achieving these milestones, we have not only laid the groundwork for future automation but also demonstrated the feasibility of employing cutting-edge computer vision techniques in real-world agricultural settings.

Regarding future next steps, we are actively exploring several avenues to enhance our pipeline's capabilities. While our current focus has been on single individuals with their entire bodies in the image due to the nature of weight prediction, we recognize the value in potentially using discarded images with multiple animals for other applications. These images could be repurposed for tasks such as animal behavior analysis or group activity monitoring, expanding the utility of our system.

Furthermore, since we have already classified images with single animals, each with its unique identification, we can leverage this data to attempt to find matching features of these specific animals within images containing multiple animals. We plan to employ techniques such as feature matching for this purpose. By doing so, we aim to expand the pool of available images for weight prediction, thereby enhancing the versatility and robustness of our system. This represents a promising research direction that we are actively exploring.

Table 2.1 Number of correct segmentation and its percentual (N=60)

| Image | Correct Segmentation | Percentual |
|---------------------|----------------------|------------|
| Depth | 43 | 72% |
| Infrared | 12 | 20% |
| Depth x Infrared ** | 54 | 90% |

** the best result.

Table 2.2 Performance of image classification using AlexNet and ResNet-50 with support vector machine (SVM) on different image types

| Type of Image ¹ | Accuracy | False Positive | False Negative | Precision | Specificity | Recall | F1 Score |
|---|----------|----------------|----------------|-----------|-------------|--------|----------|
| AlexNet | | | | | | | |
| colored depth | 0.939 | 0.104 | 0.018 | 0.904 | 0.980 | 0.982 | 0.941 |
| depth | 0.500 | 1.000 | 0.000 | 0.500 | NaN | 1.000 | 0.667 |
| infrared | 0.868 | 0.179 | 0.086 | 0.837 | 0.905 | 0.914 | 0.874 |
| combined* | 0.917 | 0.031 | 0.135 | 0.966 | 0.878 | 0.865 | 0.913 |
| ResNet-50 with support vector machine (SVM) | | | | | | | |
| colored depth | 0.770 | 0.000 | 0.460 | 1.000 | 0.685 | 0.540 | 0.701 |
| depth | 0.920 | 0.074 | 0.086 | 0.925 | 0.915 | 0.914 | 0.920 |
| infrared | 0.877 | 0.160 | 0.086 | 0.851 | 0.907 | 0.914 | 0.882 |
| combined | 0.801 | 0.043 | 0.356 | 0.938 | 0.729 | 0.644 | 0.764 |

¹Type of Image: Refers to the types of images captured by the Intel RealSense Depth Camera D435, which include depth, color, and infrared images. The 'combined' images were created by combining the colored depth and infrared images.

* the best-performing image type for image classification among the evaluated architectures.

Table 2.3 Impact of data augmentation on AlexNet Image Classification using the combined image converted to jet colormap

| Number of Images Augmented per Class ¹ | Total of Image ² | Accuracy | False Positive | False Negative | Precision | Specificity | Recall | F1 Score | Classifier Run Time (Minutes) ³ | Predicted Good Real Images ⁴ |
|---|-----------------------------|----------|----------------|----------------|-----------|-------------|--------|----------|--|---|
| Original | 1634 | 0.917 | 0.031 | 0.135 | 0.966 | 0.878 | 0.865 | 0.913 | 12.320 | 5689 |
| 100 | 1834 | 0.932 | 0.120 | 0.016 | 0.982 | 0.891 | 0.880 | 0.928 | 13.570 | 2831 |
| 200 | 2034 | 0.951 | 0.049 | 0.049 | 0.951 | 0.951 | 0.951 | 0.951 | 15.713 | 2256 |
| 300 | 2234 | 0.915 | 0.031 | 0.139 | 0.874 | 0.965 | 0.969 | 0.919 | 18.050 | 2758 |
| 500* | 2634 | 0.960 | 0.008 | 0.072 | 0.932 | 0.992 | 0.992 | 0.961 | 23.556 | 1670 |
| 700 | 3034 | 0.970 | 0.033 | 0.026 | 0.973 | 0.967 | 0.967 | 0.970 | 40.599 | 2033 |
| 900 | 3434 | 0.969 | 0.015 | 0.047 | 0.955 | 0.985 | 0.985 | 0.970 | 60.449 | 2106 |
| 1100 | 3834 | 0.973 | 0.013 | 0.042 | 0.959 | 0.987 | 0.987 | 0.973 | 70.309 | 1660 |

¹Class: Refers to the two classes of images, one containing usable images and the other containing unusable images.

²Total of Images: Refers to the total number of images across both classes. Both the usable and unusable classes have the same number of images.

³Classifier Run Time: Refers to the amount of time taken by the computer to run the image classifier network.

⁴Predicted Good Real Images: Refers to the number of good (usable) images that were correctly predicted by the trained neural network on a set of infusion images that were converted to jet colormap, excluding those used for training the network.

⁵Original Images: Refers to the original infusion images, which were converted to jet colormap images.

*500: indicates the classifier that was selected for subsequent image segmentation.

Table 2.4 Performance of AlexNet classifier trained with synthetic images on real and unseen images during training

| Number of Images Augmented per Class ¹ | Total of Image ² | Accuracy | False Positive | False Negative | Precision | Specificity | Recall | F1 Score |
|---|-----------------------------|----------|----------------|----------------|-----------|-------------|--------|----------|
| 100 | 1834 | 95.833 | 0.059 | 0.025 | 0.975 | 0.943 | 0.941 | 0.958 |
| 200 | 2034 | 95.588 | 0.020 | 0.069 | 0.935 | 0.979 | 0.980 | 0.957 |
| 300 | 2234 | 89.216 | 0.010 | 0.206 | 0.828 | 0.988 | 0.990 | 0.902 |
| 500 | 2634 | 93.382 | 0.005 | 0.127 | 0.886 | 0.994 | 0.995 | 0.938 |
| 700 | 3034 | 96.569 | 0.010 | 0.059 | 0.944 | 0.990 | 0.990 | 0.967 |
| 900 | 3434 | 96.569 | 0.000 | 0.069 | 0.936 | 1.000 | 1.000 | 0.967 |
| 1100 | 3834 | 92.892 | 0.005 | 0.137 | 0.879 | 0.994 | 0.995 | 0.933 |

¹ Class: Refers to the two classes of images, one containing usable images and the other containing unusable images.

² Total of Images: Refers to the total number of images across both classes. Both the usable and unusable classes have the same number of images.

Table 2.5 Performance of semantic segmentation on different image types

| Type of Image | Global Accuracy | Mean Accuracy | Mean IoU | Weighted IoU | Mean BFScore | False Positive | False Negative |
|---------------|-----------------|---------------|----------|--------------|--------------|----------------|----------------|
| colored depth | 0.938 | 0.897 | 0.790 | 0.892 | 0.559 | 0.159 | 0.084 |
| depth | 0.863 | 0.502 | 0.434 | 0.746 | 0.666 | 0.966 | 0.100 |
| infrared | 0.942 | 0.859 | 0.786 | 0.895 | 0.438 | 0.256 | 0.027 |
| combined | 0.918 | 0.807 | 0.716 | 0.857 | 0.484 | 0.347 | 0.040 |

Table 2.6 Performance of ResNet50 in Classifying Images of Animals: Headless Body vs. Headed Body

| Set | Accuracy | False Positive | False Negative | Precision | Specificity | Recall | F1 Score |
|------------|----------|----------------|----------------|-----------|-------------|--------|----------|
| test | 0.983 | 0.011 | 0.023 | 0.978 | 0.989 | 0.989 | 0.983 |
| validation | 0.955 | 0.023 | 0.068 | 0.935 | 0.976 | 0.977 | 0.956 |

Table 2.7 Summary of Feature Statistics for Top-View Animal Images

| Descriptive Statistics | Area | Eccentricity | Major Axes | Minor Axes | Perimeter |
|------------------------|-------|--------------|------------|------------|-----------|
| median | 45223 | 0.921 | 388.820 | 151.418 | 943.350 |
| mean | 45335 | 0.915 | 387.416 | 152.772 | 951.024 |
| standard deviation | 6284 | 0.032 | 42.187 | 11.724 | 103.630 |
| minimum | 22854 | 0.565 | 192.258 | 115.611 | 594.456 |
| maximum | 82763 | 0.958 | 560.061 | 296.267 | 1459.277 |
| lower limit | 30355 | 0.883 | 302.742 | 126.091 | 719.371 |
| upper limit | 60441 | 0.957 | 475.802 | 177.998 | 1172.835 |
| CV ¹ | 14 | 3.478 | 10.889 | 7.674 | 10.897 |

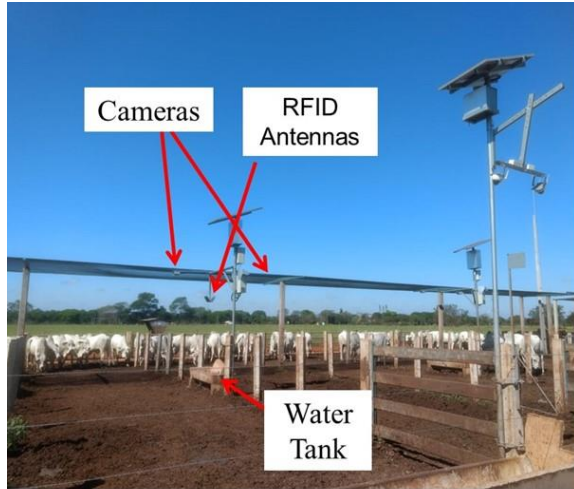
¹ CV: coefficient of variation.

Table 2.8 Intersection over Union (IoU) measures between the predicted segmented images and the ground truth segmented images

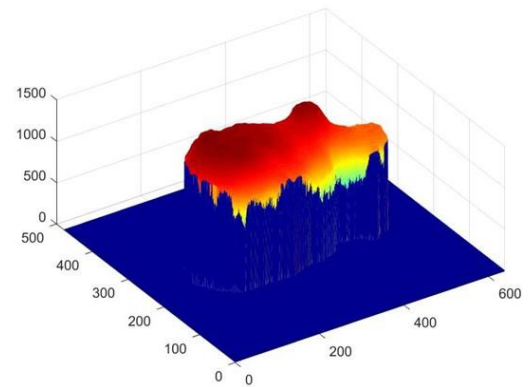
| Descriptive Statistics | IoU |
|------------------------|-------|
| mean | 0.888 |
| standard deviation | 0.024 |
| median | 0.892 |
| minimum | 0.749 |
| maximum | 0.941 |

Table 2.9 Correlation between biological features of our segmented images with biological features of gold standard images.

| Body Features | Pearson Correlation |
|---------------|---------------------|
| Area | 0.947 |
| Centroid X | 0.995 |
| Centroid Y | 0.997 |
| Major Axes | 0.802 |
| Minor Axes | 0.944 |
| Eccentricity | 0.724 |
| Orientation | 0.961 |
| EquivDiameter | 0.948 |
| Perimeter | 0.739 |
| Roundness | 0.642 |



A



B

Figure 2.1 (A) Depth image equipment consisting of RFID antennas to trigger the 3D camera when an animal approaches to the water source. (B) An example of a top-down view image collected by our equipment.

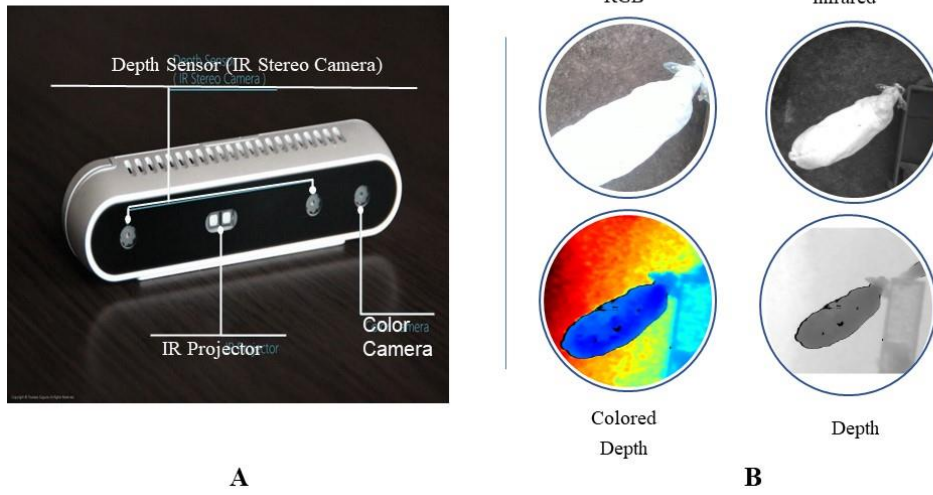


Figure 2.2 (A) The Intel® RealSense™ D435 camera used in our experiment. (B) Types of images that the 3D camera provides.

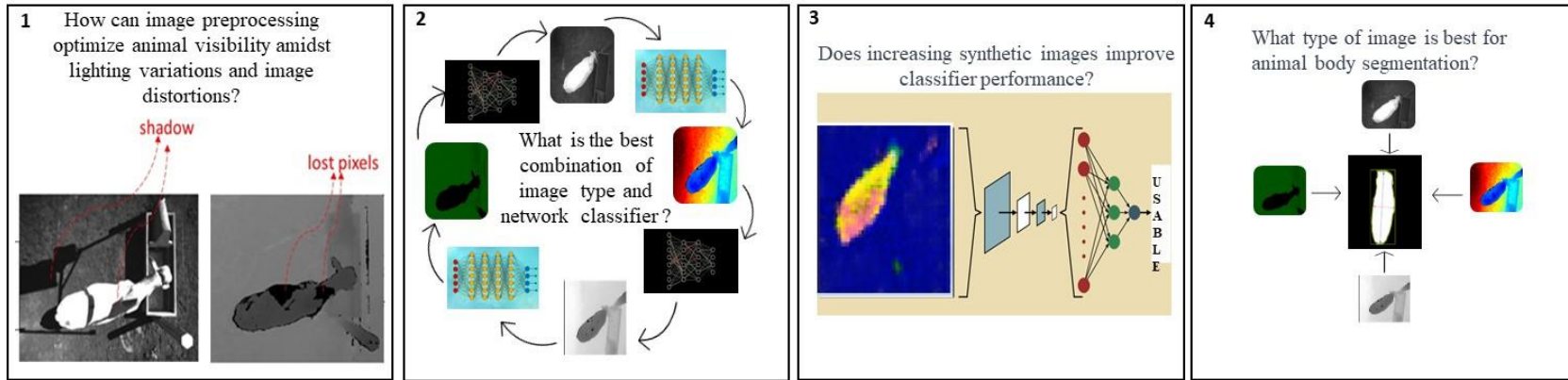


Figure 2.3 Key questions for the challenges discussed in this study.

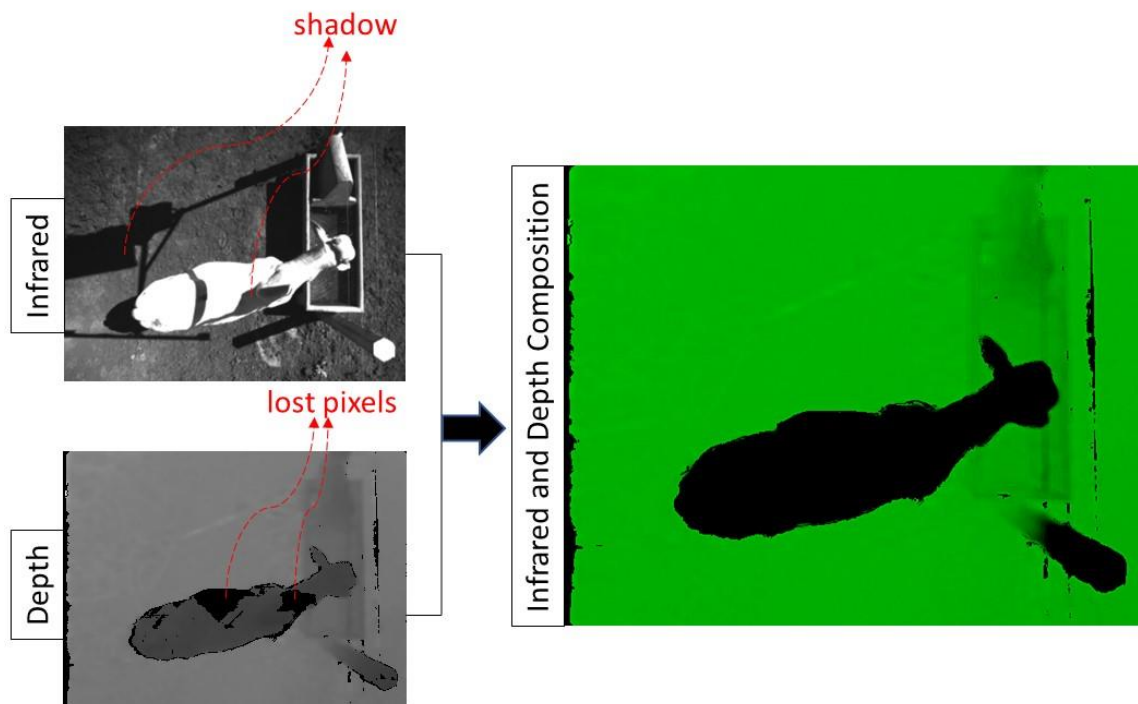
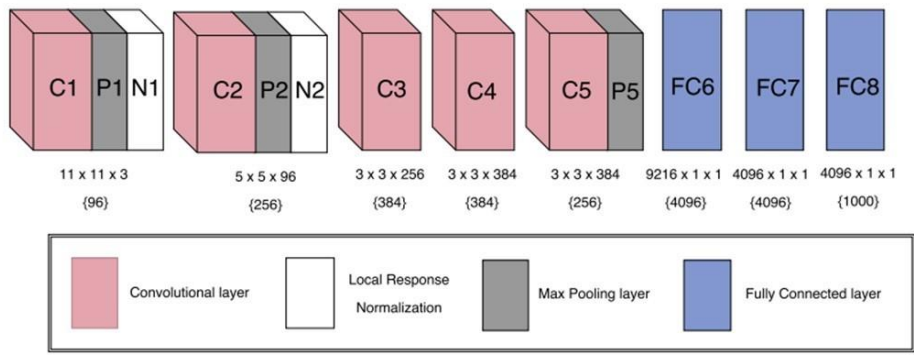
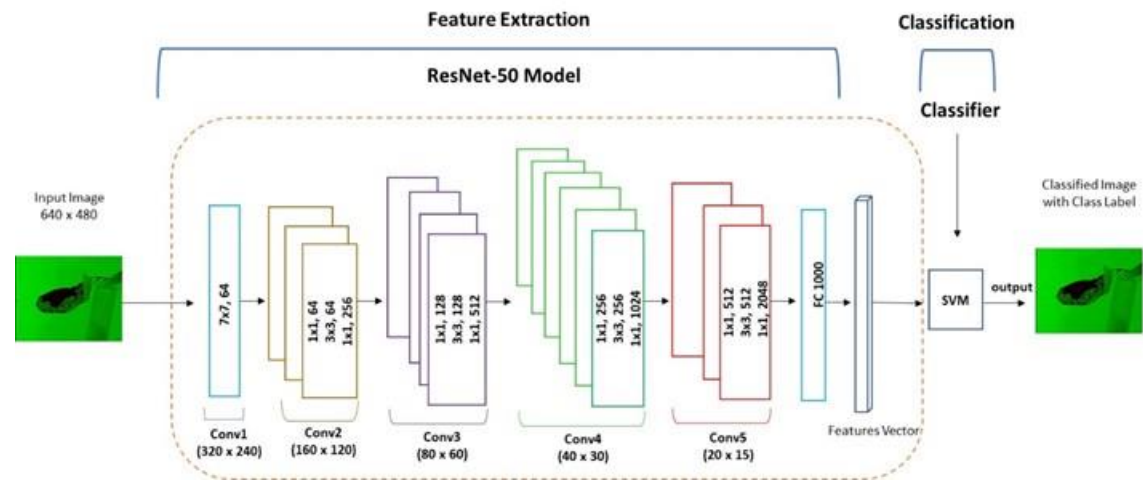


Figure 2.4 Composition of an infrared and depth images.



A



B

Figure 2.5 Flowchart depicting the selection process of the most efficient model for different combinations of neural networks and image types.

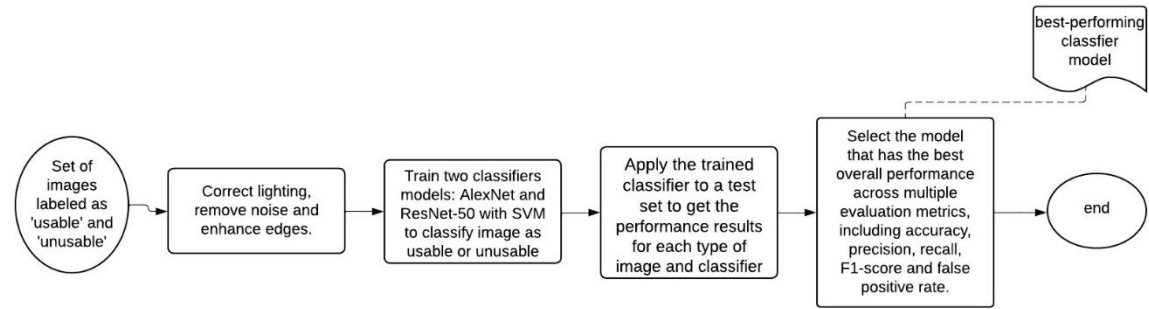
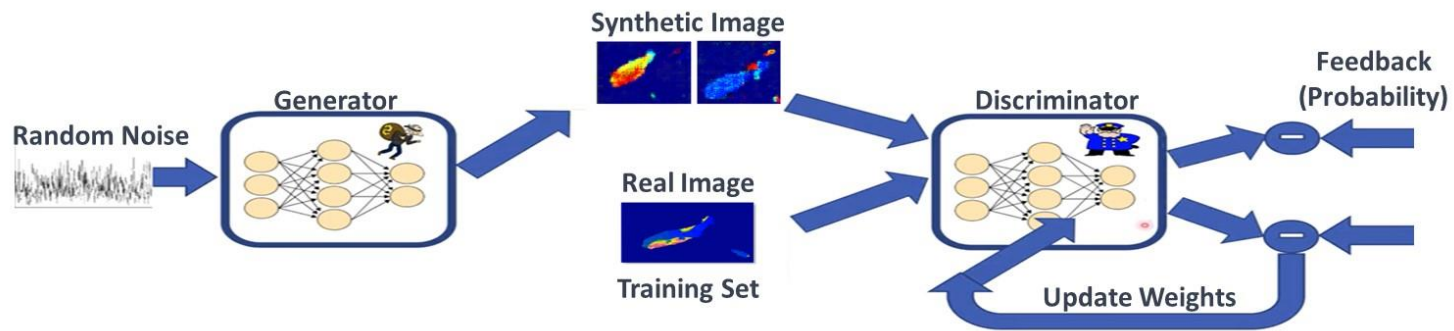
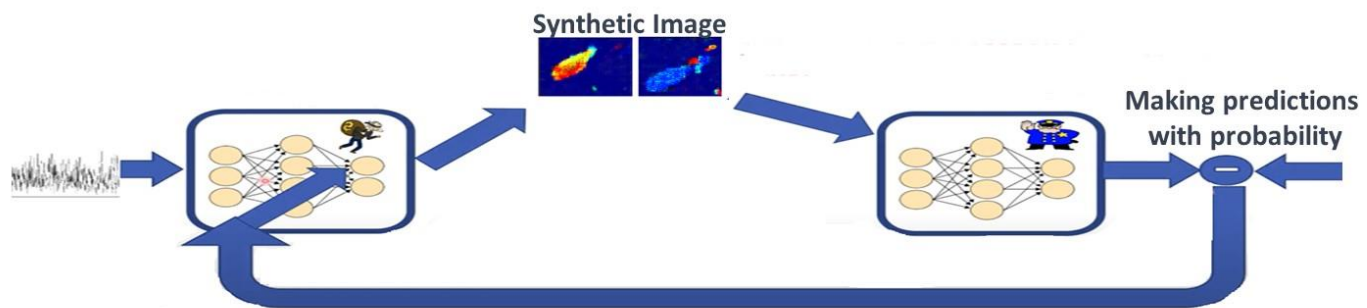


Figure 2.6 (A) AlexNet network architecture (adapted from Babu, 2016). (B) ResNet network architecture (adapted from Almagdy and Elrefaei, 2018).

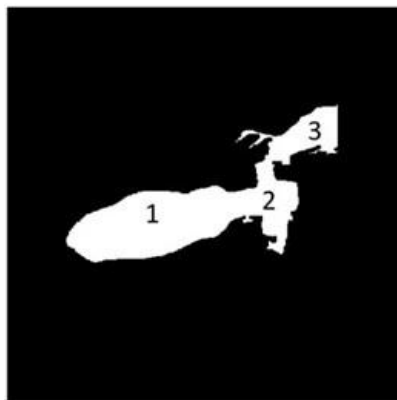


A

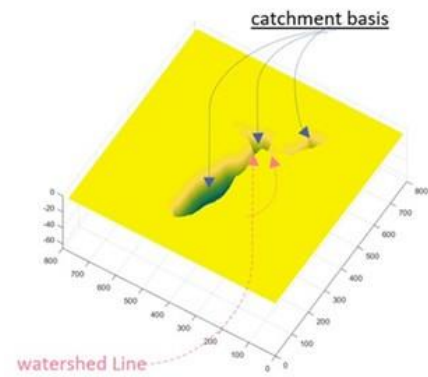


B

Figure 2.7 (A) Training of the GAN generator network. (B) Training of the GAN discriminator network.



A



B



C



D



E



F

Figure 2.8 (A) Steer binary image, and (B) Surface of the steer binary image. (C) Steer binary image, and (D) Distance transform of the steer binary image. (E) Distance transform of the steer image's complement, and (F) Negative of the distance transform of the steer image's complement.

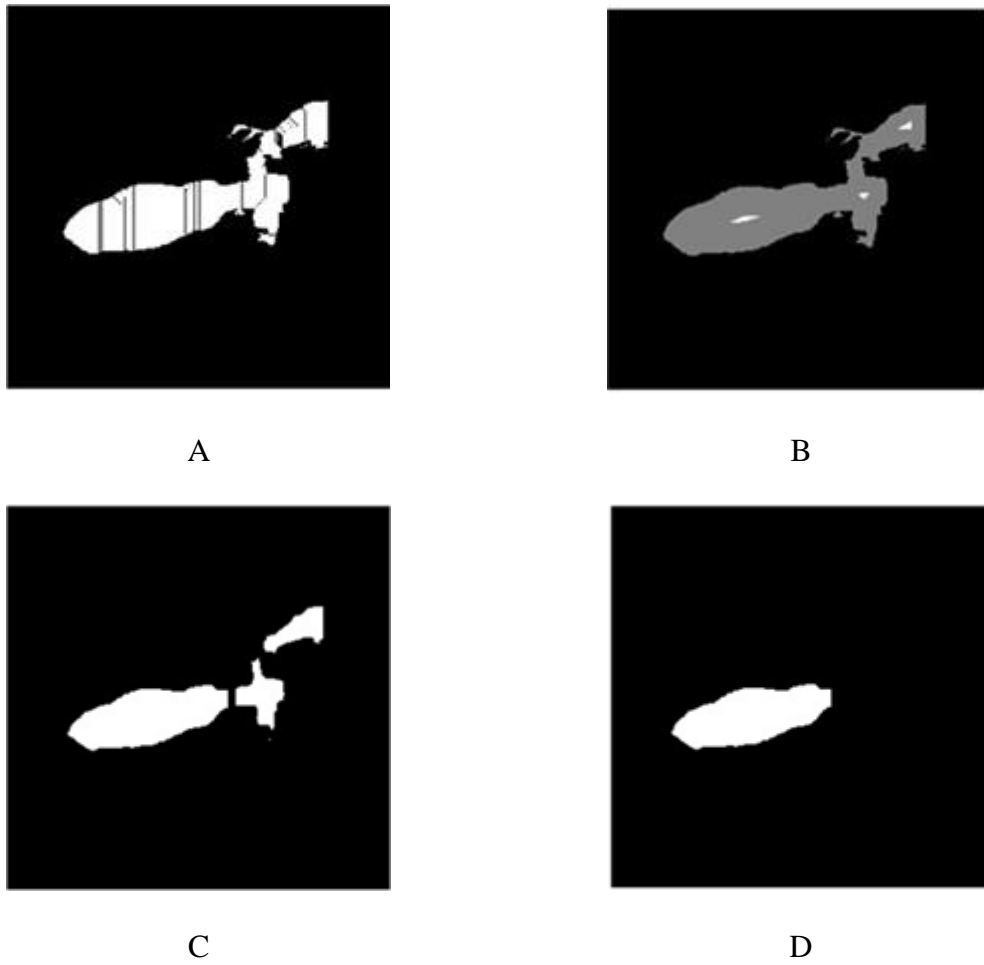
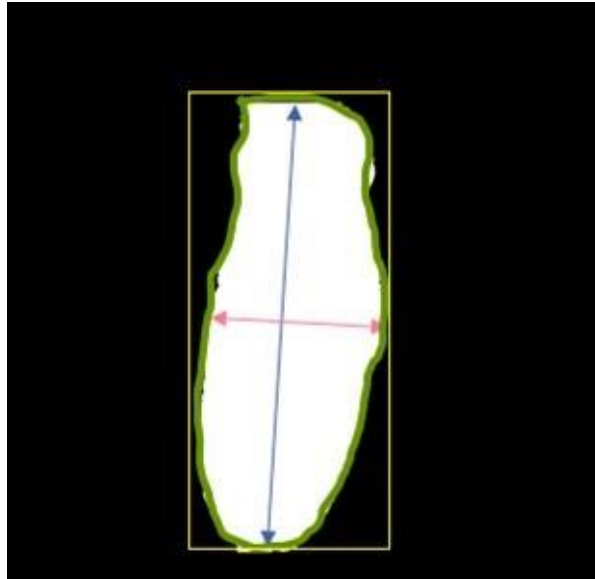


Figure 2.9 (A) Steer over-segmentation of Watershed algorithm, (B) Steer Marker-Controlled Watershed Segmentation, (C) Body separated from the other parts after Marker-Controlled Watershed Segmentation, and (D) Final segmentation – body's steer.



A

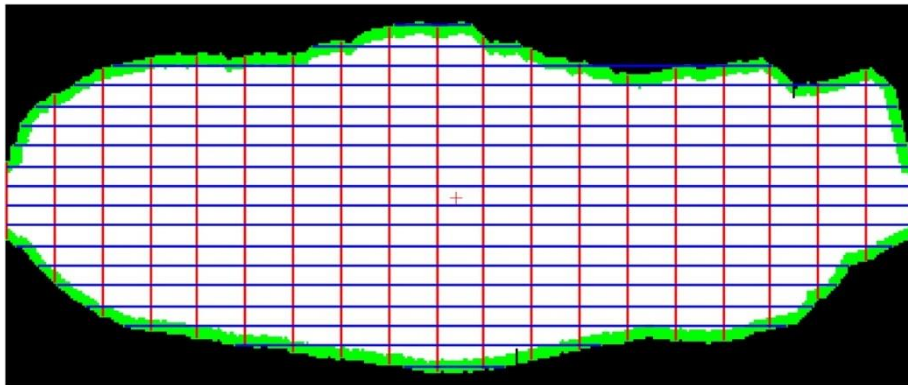


Figure 2.10 (A) Biological features from the steer. The bounding box (yellow), contour/perimeter (green), major length (blue line), minor length (red line). (B) Measuring lines for width (red line) and length (blue line).

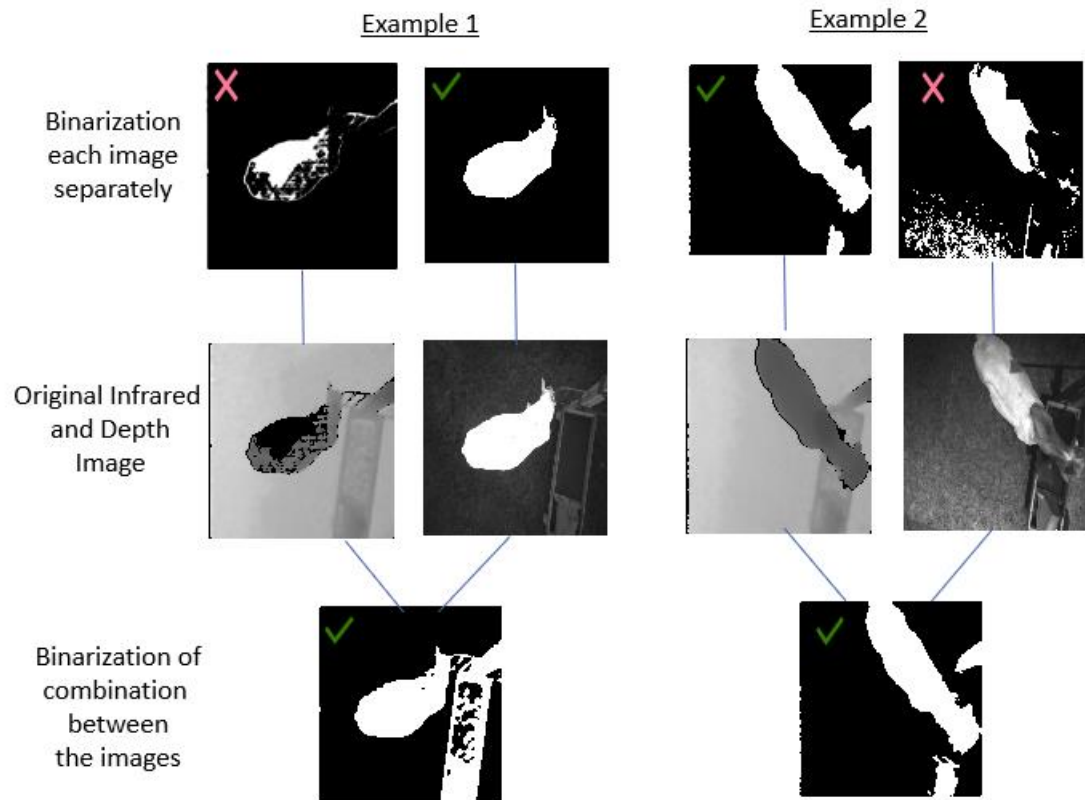


Figure 2.11 Examples of depth and infrared image segmentation and their respective combinations.

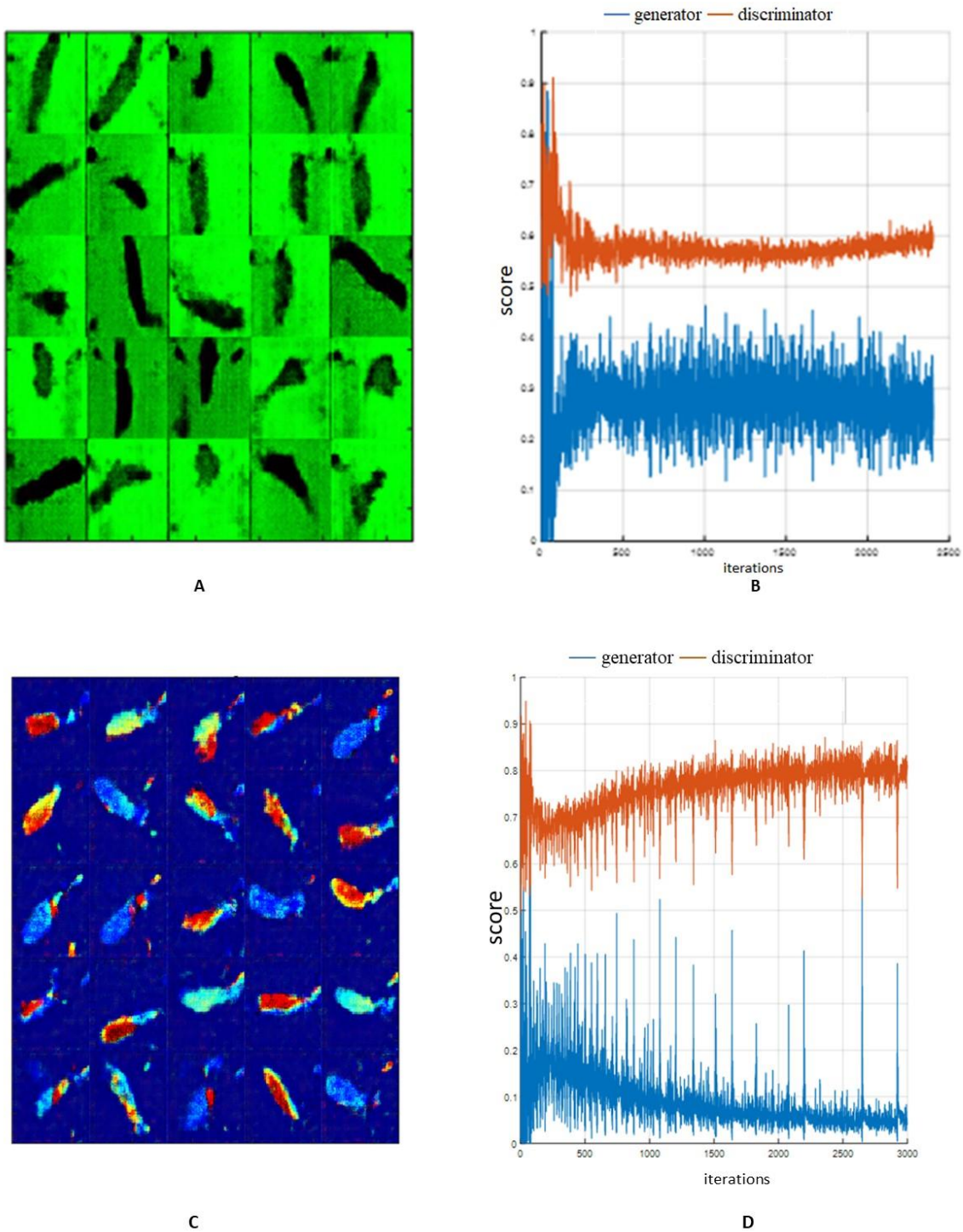
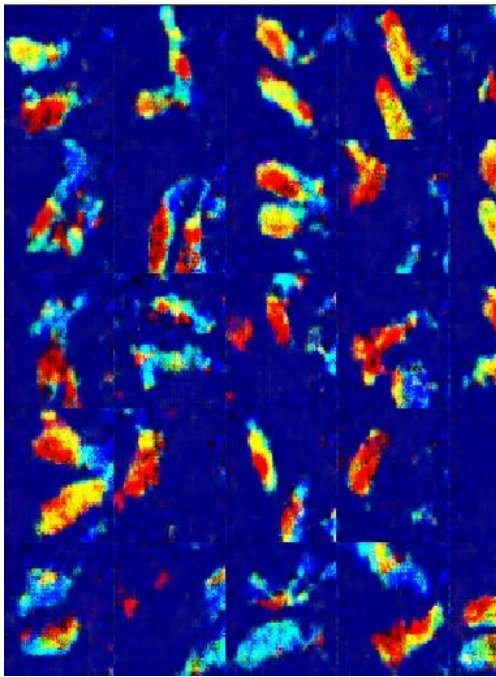
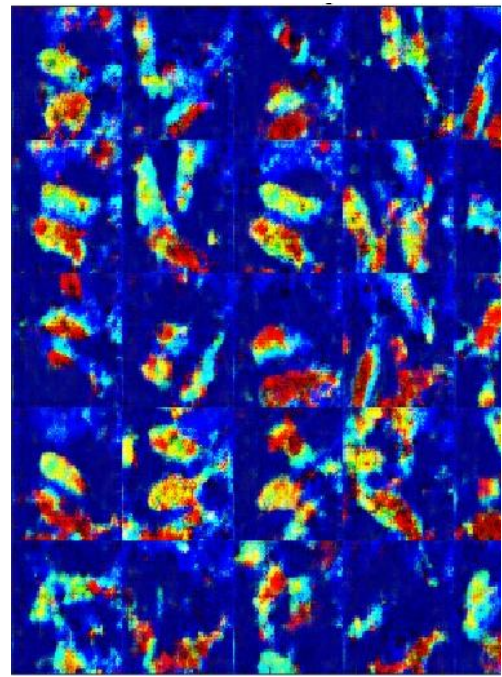


Figure 2.12 Evaluation of GAN model performance. (A) usable synthetic images generated by the GAN model using the combined image. (B) Score of generated images by the generator (green) and discriminator (orange) over interactions. (C) usable synthetic images generated by the GAN model using colorjet image. (D) Score of generated images by the generator (green) and discriminator (orange) over interactions.



A



B

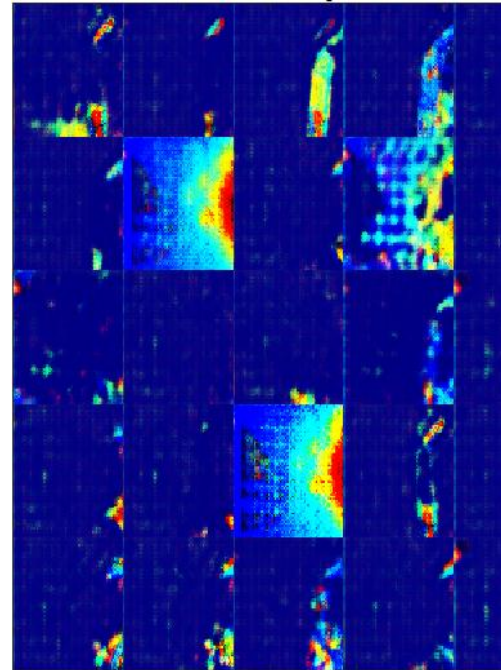
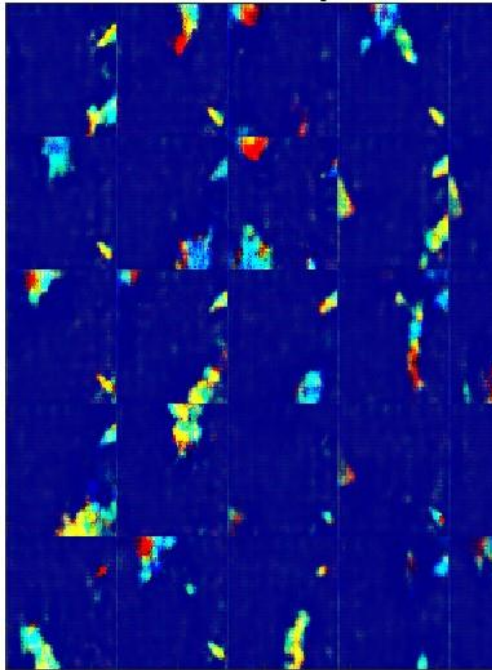


Figure 2.13 Sample of 25 synthetic images generated from the colorjet image, including four categories of unusable images. (A) Two animals. (B) Multiple animals. (C) Partial views of one animal. (D) background or image error.

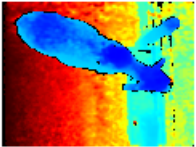


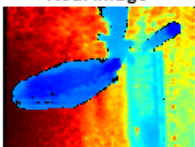






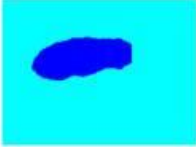



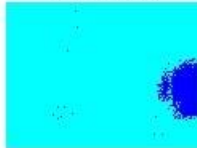




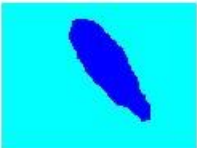

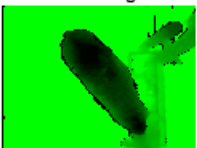

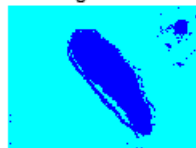
| Type | The worst Segmentation Case | | | The best Segmentation Case | | |
|---------------|---|---|---|---|---|--|
| colored depth | Real Image  | Real Label  | Segmented  | Real Image  | Real Label  | Segmented  |
| depth | Real Image  | Real Label  | Segmented  | Real Image  | Real Label  | Segmented  |
| infrared | Real Image  | Real Label  | Segmented  | Real Image  | Real Label  | Segmented  |
| combined | Real Image  | Real Label  | Segmented  | Real Image  | Real Label  | Segmented  |

Figure 2.14 The worst and the best semantic segmentation case for each type of image.

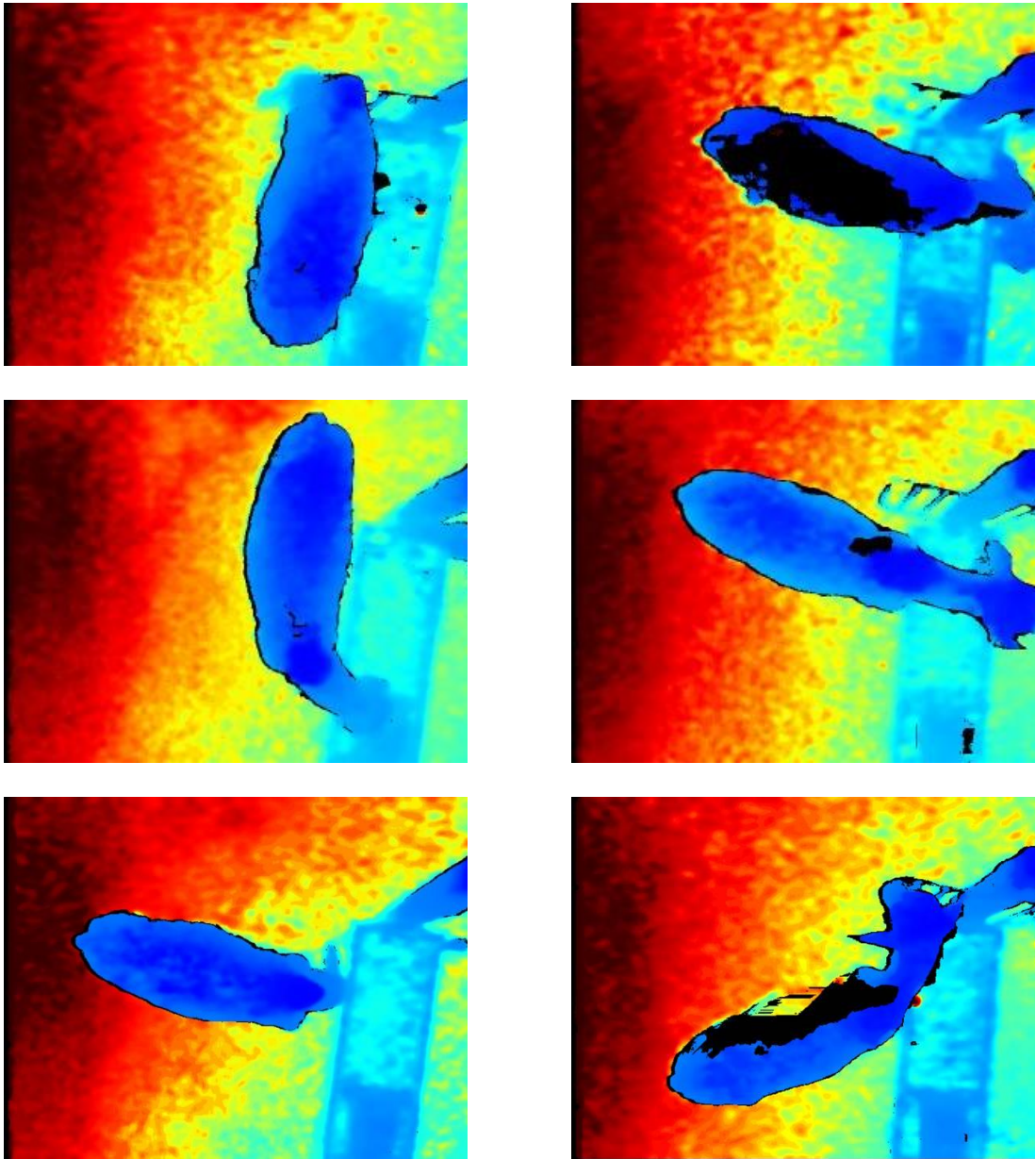


Figure 2.15 Top-view images showcasing feedlot cattle image diversity in terms of position, illumination, and lack of pixels in the images.

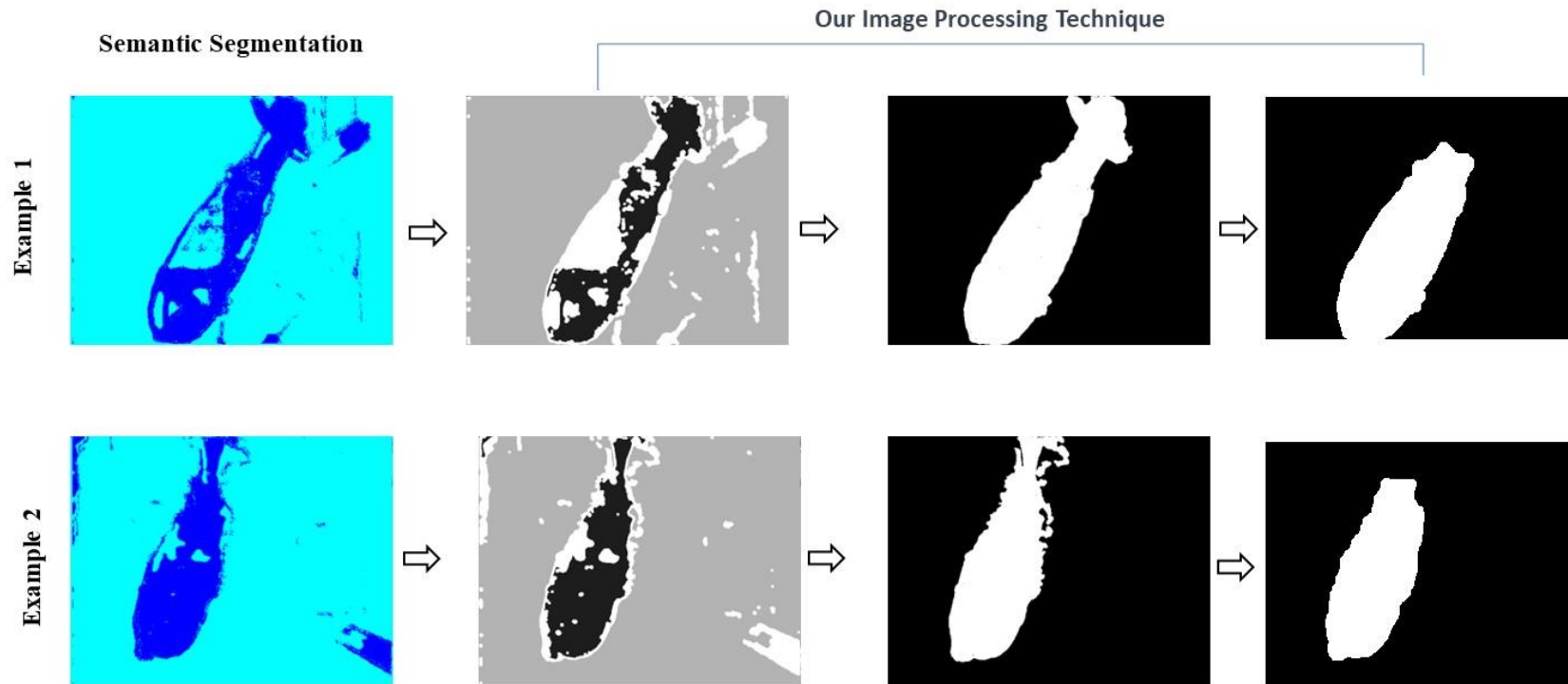


Figure 2.16 Examples of the output for semantic segmentation and our image processing technique for improving its image quality.

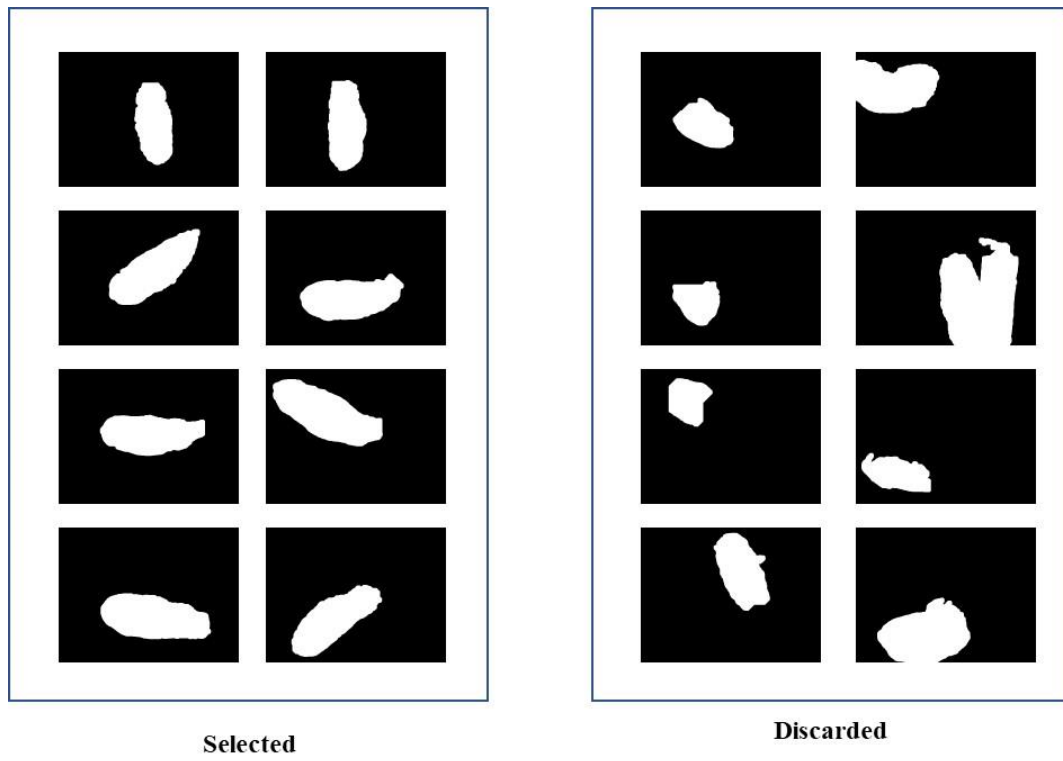


Figure 2.17 Examples of selected and discarded images in classifier-based filtering and false positive reduction.

References

- Alhwarin, F., A. Ferrein, and I. Scholl. 2014. IR Stereo Kinect: Improving depth images by combining structured light with IR stereo. *Proc. PRICAI 2014: Trends in Artificial Intelligence. Lecture Notes in Computer Science*. 8862:409–421. doi:10.1007/978-3-319-13560-1_33.
- Almabdy, S., and L. Elrefaei. 2019. Deep convolutional neural network-based approaches for face recognition. *Appl. Sci.* 9:4397. doi:10.3390/app9204397.
- Alsabei, A., A. Alsayed, M. Alzahrani, and S. Al-Shareef. 2021. Waste classification by Fine-Tuning Pre-trained CNN and GAN. *Int. J. Comput. Sci. Netw. Secur.* 21:65–70. doi:10.22937/IJCSNS.2021.21.8.9.
- Asano, H., E. Hirakawa, H. Hayashi, K. Hamada, Y. Asayama, M. Oohashi, A. Uchiyama, and T. Higashino. 2022. A method for improving semantic segmentation using thermographic images in infants. *BMC Med. Imaging* 22:1. doi:10.1186/s12880-021-00730-0.
- Bargshady, G., X. Zhou, P.D. Barua, R. Gururajan, Y. Li, and U. R. Acharya. 2022. Application of CycleGAN and transfer learning techniques for automated detection of COVID-19 using X-ray images. *Pattern Recognit. Lett.* 153:67–74. doi:10.1016/j.patrec.2021.11.020.
- Beucher, S. 2000. The Watershed Transformation Applied to Image Segmentation. *Scanning Microsc.* 1992:28.
- Bird J. J., C. M. Barnes, L. J. Manso, A. Ekárt, and D. R Faria. 2022. Fruit quality and defect image classification with conditional gan data augmentation. *Sci. Hortic.* 293:110684. doi:10.1016/j.scienta.2021.110684.
- Buecher, S., and C. Lantuéjoul. 1979. Use of watershed in contour detection. *Proc. International Workshop on Image Processing, Real-Time Edge and Motion Detection/Estimation.* 17:21.
- Chatterjee, S., D. Hazra, Y. C. Byun, and Y. W. Kim. 2022. Enhancement of image classification using transfer learning and GAN-based synthetic data augmentation. *Mathematics* 10:1541. doi:10.3390/math10091541.
- Coates, A., H. Lee, and A. Y. Ng. 2011. An analysis of single-layer networks in unsupervised feature learning. *Proc. 14th International Conference on AI and Statistics.* 215–223.
- Cominotte, A., A. F. A. Fernandes, J. R. R. Dorea, G. J. M. Rosa, M. M. Ladeira, E. H.C.B. Van Cleef, G. L. Pereira, W. A. Baldassini, and O. R. Machado Neto. 2020. Automated computer vision system to predict body weight and average daily gain in beef cattle during

- growing and finishing phases. *Livest Sci.* 232:103904–14. doi:10.1016/j.livsci.2019.103904.
- Deng, J., W. Dong, R. Socher and L.-J Li. 2009. Imagenet: A large-scale hierarchical image database. *Proc. Computer Vision and Pattern Recognition.* 248-255.
- Digabel, H., and C. Lantuejoul. 1978. *Proc. Actes du Second Symposium Europeen de Analyse Quantitative des Microstructures en Sciences des Materiaux.* Biol. Med. 1:85–99.
- Fernandes, A. F. A., Dórea, J. R. R., Fitzgerald, R., Herring, W., and Rosa, G. J. M. 2019. A novel automated system to acquire biometric and morphological measurements and predict body weight of pigs via 3D computer vision. *J. Anim. Sci.* 97:496–508. doi:10.1093/jas/sky418.
- Frid-Adar, M., I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan. 2018. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing.* 2018: 321–331. doi:10.1016/j.neucom.2018.09.013.
- Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. 2014. Generative adversarial nets. *Proc. Conference on Advances in Neural Information Processing Systems.* 1:2672–2680.
- Gomez, W., L. Leija, A. V. Alvarenga, A. F. C. Infantosi, and W. C. A. Pereira. 2010. Computerized lesion segmentation of breast ultrasound based on marker-controlled watershed transformation. *Med Phys.* 37:82-95. doi:10.1118/1.3265959.
- Gupta, D., B. Khatri, G. Dhiman, M. Soni, S. Gomathi, and D. Mane. 2021. Review of ECG arrhythmia classification using deep neural network. *Mater. Today Proc.* 43:518–524. doi:10.1016/j.matpr.2021.05.249.
- Gunilla, B. 1986. Distance transformations in digital images. *Computer Vision, Graphics, and Image Processing. Computer vision, graphics, and image processing.* 34:344–371. doi:10.1016/S0734-189X(86)80047-0.
- Hazra, D., Y. C. Byun, W. J. Kim, and C. U. Kang. 2022. Synthesis of microscopic cell images obtained from bone marrow aspirate smears through Generative Adversarial Networks. *Biology.* 11:276. doi: 10.3390/biology11020276
- He, K., X. Zhang, S. Ren, and J. Sun 2016a. Deep residual learning for image recognition. *Proc. IEEE Conference on Computer Vision and Pattern Recognition* 1:770–778.
- He, K., X. Zhang, S. Ren, and J. Sun 2016b. *Proc. Identity mappings in deep residual networks. Proc. European Conference on Computer Vision* 1:630–645.

- Kadlec, R., S. Indest, K. Castro, S. Waqar, L. M. Campos, S. T. Amorim, Y. Bi, M. D. Hanigan, and G. Morota. 2022. Automated acquisition of top-view dairy cow depth image data using an RGB-D sensor camera. *Transl. Anim. Sci.* 6:4. doi:10.1093/tas/txac163.
- Krizhevsky, A., I. Sutskever, and E. H. Geoffrey. 2017. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM.* 60: 84–90. doi: 10.1145/3065386.
- Kwasniewska, A., J. Ruminski, J. and M. Szankin. 2019. Improving accuracy of contactless respiratory rate estimation by enhancing thermal sequences with deep Neural Networks. *Applied Sciences* 9:4405. doi:10.3390/app9204405.
- Long, J., E. Shelhamer, and T. Darrell. 2015. Fully Convolutional Networks for Semantic Segmentation. *Proc. Conference on Computer Vision and Pattern Recognition* 3431-3440. doi: 10.1109/CVPR.2015.7298965.
- Lu, Y., D. Chen, E. Olaniyi, and Y. Huang. 2022. Generative adversarial networks (GANs) for image augmentation in agriculture: A systematic review. *Comput. Electron. Agric.* 200:107208. doi:10.1016/j.compag.2022.107208.
- Lux, F., and P. Matula. 2019. DIC Image Segmentation of Dense Cell Populations by Combining Deep Learning and Watershed. *Proc. IEEE 16th International Symposium on Biomedical Imaging.* 1:236–239.
- MATLAB R2022a. (Version 10.2.1). MathWorks. Retrieved from <https://www.mathworks.com/products/matlab.html>.
- Meyer, F., and S. Beucher. 1990. Morphological segmentation. *Proc. Journal of Visual Communication and Image representation.* .1:21-46.
- Otsu, N. 1979. A Threshold Selection Method from Gray-Level Histograms. *Proc. IEEE Transactions on Systems.* 9:62–66.
- Quoc, L., J. Ngiam., Z. Chen, D. Chia, P.W. Koh, and A. Ng. 2010. Tiled convolutional neural networks. *Proc. Neural Information Processing Systems.*
- Shen, X., Ma, H., Liu, R., Li, H., He, J., and Wu, X. 2021. Lesion segmentation in breast ultrasound images using the optimized marked watershed method. *BioMed Eng OnLine.* 20:57. doi:10.1186/s12938-021-00891-7.
- Srinivas, S., R. K. Sarvadevabhatla, K. R. Mopuri, N. Prabhu, S. S. S. Kruthiventi, and R. Venkatesh Babu. 2016. A taxonomy of deep convolutional neural nets for computer vision. *Front. Robot. AI.* 2:36. doi:10.3389/frobt.2015.00036.
- Srivastav, D., A. Bajpai, and P. Srivastava. Improved classification for pneumonia detection using transfer learning with GAN based synthetic image augmentation. 2021. *Proc. 11th*

- International Conference on Cloud Computing, Data Science and Engineering.11:433–437.doi: 10.1109/Confluence51648.2021.9377062.
- Thamaraimanalan, T., and P. Sampath. 2019. A low power fuzzy logic-based variable resolution ADC for wireless ECG monitoring systems. *Cogn. Syst. Res.* 57:236–245. doi:10.1016/j.cogsys.2018.10.033.
- Xu, S., H. Liu, and E. Song. 2011. Marker-controlled watershed for lesion segmentation in mammograms. *J. Digit. Imaging.* 24:754-763. doi:10.1007/s10278-011-9365-2.
- Ye, D., and R. Han. 2021. Image semantic segmentation method based on improved ERFNet model. *J. Eng.* 2022:180-190. doi:10.1049/tje2.12104.
- Zak, J., M. K. Grzeszczyk, A. Pater, L. Roszkowiak, K. Siemion, and A. Korzynska. 2022. Cell image augmentation for classification task using GANs on pap smear dataset. *Biocybern. Biomed. Eng.* 42:995-1011. doi:10.1016/j.bbe.2022.07.003.
- Zhao, W., Y. Fu, X. Wei, and H. Wang. 2018. An improved image semantic segmentation method based on superpixels and conditional random fields. *Appl. Sci.* 8:837. doi:10.3390/app8050837.
- Zhong, J., M. Li, X. Liao, and J. A. Qin. 2020. Real-Time infrared stereo matching algorithm for RGB-D cameras' indoor 3D perception. *ISPRS Int. J. Geo-Inf.* 9:472. doi:10.3390/ijgi9080472.

CHAPTER THREE: COMPUTER VISION SYSTEM FOR BEEF CATTLE WEIGHT
PREDICTION IN COMMERCIAL FEEDLOTS: INNOVATIONS IN IMAGE FEATURE
EXTRACTION

Abstract

The goal of this study was to develop a computer vision system for body weight (BW) prediction of beef cattle utilizing advanced image feature extraction techniques on top-down images in outdoor feedlot settings. Our research conducted a comparative analysis between conventional biological feature extraction methods from animal body images, such as area, perimeter, width, and length, and Fourier features obtained through the application of Fourier transformation to the images. The study also compared various machine learning regression techniques for the prediction of animals' BW using the biological and Fourier features as predictors, including Partial Least Squares Regression, Ridge Regression, Lasso Regression, and Artificial Neural Networks (ANN). The results underscored the pivotal significance of gathering more comprehensive features due to the inherent variability in the animals' positions within these dynamic environments. Fourier transformation consistently outperformed traditional biological features across all evaluated models. The combination of Fourier transformation with ANN significantly outperformed other modeling strategies, resulting in Mean Bias about 4.3 times lower, root mean squared error (RMSE) about 3.2 times lower, and mean absolute error (MAE) about 5.3 times lower. Furthermore, the correlation between predicted and observed weights reached a notable 0.98, surpassing the biological model's 0.77. On average, the ANN, when harnessed with Fourier features, attained predictions approximately three to four times more accurate than those obtained through other models. These findings unequivocally advocated the ANN integrated with Fourier

transformation as the most promising approach for efficient prediction of beef cattle BW in outdoor settings.

Introduction

The feedlot sector represents an intensive production system with the aim of raising and fattening cattle until they reach the appropriate slaughter weight. The possibility of accurately assessing weight and development of animals in beef feedlot operations would certainly help on multiple fronts, including the potential of early disease detection, optimizing feed allocation, and determining the ideal slaughter point (Ulutas et al., 2002). This would not only enable cattle producers to adopt proactive disease management strategies, intervening promptly and minimizing the economic impact of illnesses but also facilitate optimizing feed management, avoiding waste, and improving feed efficiency. Furthermore, accuracy in determining the appropriate slaughter time based on live weight ensures carcass quality and better economic return (Heinrichs et al., 1992).

Traditionally, the assessment of the body weight (BW) of beef cattle has been carried out using weighing scales. However, bringing animals periodically to a central scale is impractical due to unbearable stress to the animals and significant human labor costs and logistics (Schofield et al., 1999; Cooke, 2014; Haskell et al., 2014; Lees et al., 2020). Alternatively, several technology companies currently offer systems that passively capture total or partial body weight (Wang et al., 2021), which can be allocated within each feedlot pen. Nevertheless, these technologies, in addition to suffering from issues related to missing data records, require frequent maintenance, and incur considerable costs, potentially affecting the final profits of feedlot operations.

In response to these challenges, computer vision systems (CVS) have emerged as an effective and non-invasive tool for predicting BW (Fernandes et al., 2019; Fernandes et al., 2020; Cominotte et al., 2020; Oliveira et al., 2021; Dorea & Rosa, 2022; Hou et al., 2023). These seminal works collectively underscore the significant potential of image-based predictions of live weight, solidifying the notion that precise weight prediction using images is indeed attainable in the field of livestock management. Most of these advances are reviewed in Dohmen et al. (2020), which primarily focuses on the analysis of images collected with animals held in restricted locations for image acquisition, highlighting the effectiveness of basic attributes of body shape and size such as area, withers height, hip height, body length, hip width, body volume, and chest circumference in weight estimation.

While these groundbreaking studies have laid a solid foundation for the successful application of CVS in BW estimation of livestock, primarily in indoor conditions, our research endeavors to extend the methodology to real-world feedlot environments (Padilha et al., 2023). In our study, we captured images in the authentic outdoor environment of feedlots with unrestrained animals, encountering challenges such as variations in animal positions over time, fluctuations in ambient light and temperature, and other complexities. Such variations, when not appropriately addressed, can impair the accuracy of image-derived features and consequently affect the precision of BW predictions. In this context, we explored techniques to decompose images into their individual signals, with the expectation of extracting more detailed image characteristics and thus enhancing the quality of BW predictions.

To achieve this goal, we investigated the use of Fourier Transform, as suggested for example by Guesmi et al. (2015), Mironovova et al. (2015), Hassen et al. (2014), and Faraggi (2019). This technique has demonstrated ability to capture oscillations and periodic variations in

images, phenomena that may not be easily identified through traditional measures. The application of Fourier transforms has the potential to enable a more in-depth analysis of images, thus contributing to improving the quality of BW predictions.

In summary, the objective of this article was twofold. First, we conducted a comprehensive and comparative evaluation of the effectiveness of Fourier feature extraction techniques relative to traditional image-derived features. Second, we investigated and compared four machine learning techniques (partial least squares regression, ridge regression, lasso regression, and artificial neural networks) in terms of their prediction quality for BW of beef cattle in commercial feedlot environments. These two components together contribute to the development of an automated computer vision system to monitor BW and growth, thereby enhancing the efficiency and quality of beef cattle finishing phase.

Materials and Methods

In this section we provided an overview of our dataset, data analysis strategy, and the specific methodologies employed, encompassing Fourier Transform, Ridge Regression, Lasso Regression, Partial Least Squares Regression, and Artificial Neural Network.

Dataset

The data set was collected at DSM Innovation and Applied Center, in Brazil. Top-view images from 3D cameras (Intel® RealSense™ Depth Camera D435) were captured when animals approached the water tank, and their unique animal identification numbers (electronic ear tags) were recorded. Simultaneously, the animals were weighed using passive scales. For further details regarding the experimental setup, image processing techniques, and the data selection process,

refer to our previous publication, Padilha et al. (2023). The image collection and body weight measurement occurred entirely passively and independently from each other. This means that the two systems, cameras and scales, were placed in different locations within each pen, and that they were activated whenever an animal approached them. As such, by chance an animal may end up with multiple images on a single day, but without a reliable body weight measurement or, contrary, an animal may have a body weight measurement without any good quality image close enough in time to it. For the purpose of this study, we selected only those sets of images and body weights from the same animals obtained in consecutive days, resulting in 882 images with recorded weights for 79 individual animals. The descriptive statistics for body weight are presented in the Table 3.1 and the final number of images per animal in Figure 3.1.

Data analysis

The primary objective of this study was to estimate live weight of cattle in feedlots using computer vision. For such, we extracted features from top-down images of beef cattle, which served as predictors in regression models. The extracted features fall into two categories: 1) Biological Features, encompassing 49 characteristics like Area, Major Axis, Minor Axis, Eccentricity, Circularity, Equivalent Diameter, Solidity, Perimeter, Roundness, as well as 20 width and 20 length measurements, obtained using the 'regionprops' function in MATLAB (Release 2021b); and 2) Fourier Features, calculated by applying Fourier Transform to binary images of the animals' bodies using the 'fft2' function in MATLAB R2022a. These Fourier features are further dimensionally reduced using Principal Component Analysis (PCA) to retain components that account for at least 95% of the total data variation. For the biological features, as a first step of

descriptive analysis, the Pearson correlation coefficient was computed between the animal BW and each of the features extracted from the animal's body.

For the BW prediction, four distinct machine learning regression techniques, specifically Partial Least Squares Regression (PLS), Ridge Regression (RR), Lasso Regression (LR), and Artificial Neural Network (ANN), were evaluated in terms of their prediction quality. The comparative analysis focuses on assessing the efficacy of the two feature extraction methods within each of the prediction algorithms. In all cases, the predictive quality was assessed using a leave-one-out cross-validation (LOOCV) approach. A comprehensive set of evaluation metrics was employed, including mean bias, root mean squared error (RMSE), mean squared percentage error (RMSPE), mean absolute error (MAE), and the correlation between predictions and actual weights. Additionally, we estimated the 95% confidence interval to provide a comprehensive assessment of the predictive performance of each approach and model. Below, we provide some additional details on the Fourier Transform methodology as well as the prediction algorithms, i.e. Partial Least Squares Regression, Ridge Regression, Lasso Regression, and Artificial Neural Network.

Fourier Transform

The Fourier Transform (Cooley and Tukey, 1965) is a technique that allows the representation of an image as a combination of complex waves, each with varying sizes, frequencies, and phases. To illustrate the concept, consider the application of the Fourier Transform to a binary image, such as a simple square (Figure 3.2). This enables to decompose the image in terms of wave patterns. Applying the Fourier Transform to this binary image reveals two primary waves: one horizontal and one vertical. These waves emerge due to the horizontal and

vertical edges of the square. Essentially, the horizontal and vertical waves represent the fundamental components of the binary square image, each with its unique frequency (how many times the wave repeats) and phase (where it starts).

In this way, the Fourier Transform aids in understanding how an image is constructed from these basic components, i.e., the waves. In simple terms, the direction of these waves are directly related to the edge pattern and geometry of the picture. Applying this concept to the task of extracting features from an animal's body, as shown in the binary image in Figure 3.3, allows us to use these primary waves to capture the essence of the animal's geometric formation. Mathematically, in the context of images, we refer to this process as the Discrete Fourier Transform (DFT). In this case, the DFT enables the analysis of an image in terms of the contributions of different frequencies and patterns present within it. Its formula can be described as:

$$F(k, l) = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} f(i, j) e^{-l2\pi\left(\frac{ki}{M} + \frac{lj}{N}\right)},$$

in which the coefficient $F(k, l)$ is computed for each pair of values (k, l) , with the term $e^{-l2\pi\left(\frac{ki}{M} + \frac{lj}{N}\right)}$ representing a sinusoidal wave in the Fourier space. Each pair (k, l) corresponds to a specific frequency and phase of this wave, and the DFT performs the decomposition of the image into a series of sinusoidal waves, each with its unique frequency and phase. It is worth noting that the same way one can transform an image from the spatial domain to the Fourier domain using the Fourier Transform, one can also reverse this process and transform an image from the Fourier domain back to the spatial domain using the so-called "Inverse Fourier Transform."

The Fourier Transform produces a complex-valued output image that can be displayed in two ways: with real and imaginary parts or with magnitude and phase. In image processing, it is common to present the magnitude (e.g., Figure 3.3), as it contains most of the information about

the spatial domain image's structure. However, to accurately revert the Fourier image to the spatial domain after some processing in the frequency domain, it is crucial to preserve both magnitude and phase information of the Fourier image.

Ridge and Lasso Regression

In Ridge Regression and Lasso Regression, a penalty or constraint term is added to the residual sum of squares, i.e. the sum of squared prediction errors. This penalty (λ) helps to prevent overfitting and reduce the complexity of the model by controlling the magnitude of the coefficients. In Ridge Regression (Hoerl and Kennard, 1970a,b), λ penalizes coefficients with very high values, reducing their influence on the regression function. It acts as a smooth control, leading to a gradual decrease in the coefficient values. The estimate of the parameters in Ridge Regression, denoted as $\hat{\beta}^{Ridge}$, is obtained by minimizing the following objective function:

$$\hat{\beta}^{Ridge} = \arg \min_{\beta} \left\{ RSS + \lambda \sum_{j=1}^k \beta_j^2 \right\}$$

where $RSS = \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^k \beta_j X_{ij})^2$ is the residual sum of squares.

In Lasso Regression (Tibshirani, 1996), λ has the special property of setting some of the coefficients exactly to zero. This enables automatic variable selection by eliminating less relevant variables from the model. Lasso tends to generate more sparse models with a smaller number of significant predictor variables. The estimate of the parameters in Lasso Regression, denoted as $\hat{\beta}^{Lasso}$, is obtained by minimizing the following objective function:

$$\hat{\beta}^{Lasso} = \arg \min_{\beta} \left\{ RSS + \lambda \sum_{j=1}^k |\beta_j| \right\}$$

Both ridge and lasso regression put a constraint on the coefficients by introducing a penalty factor. However, while ridge regression takes the square (i.e., L2 regularization), lasso regression takes the magnitude of the coefficients (i.e., L1 regularization). The Ridge and Lasso regression models were trained using the "glmnet" function with the training data, and the optimal value of lambda was determined using the cross-validation method "cv.glmnet".

Partial Least Squares

Partial least squares regression (PLSR) is a linear regression method that applies principles similar to PCA. The distinction in a regression context is that we work with two matrices: one containing predictor variables (X) and the other containing response variables (Y). In PLSR, we decompose data using latent variables, computing scores (T) and loadings (P) to understand the relationship between X and Y. The score matrix (T) captures the underlying patterns in X, while the loading matrix (P) describes how these patterns relate to X. Additionally, another loading matrix (C) is used to relate the patterns captured by T to Y. This relationship is expressed by the equations: $X = T \cdot P'$ and $Y = T \cdot C'$. This method is particularly useful when dealing with complex variable interactions.

In essence, PLSR helps us find these latent vectors to understand the relationship between X and Y. It's particularly useful when dealing with datasets where there might be complex interactions among X. In our study, we used PLSR and determined the optimal number of components based on model performance, specifically using Root Mean Squared Error (RMSE) through cross-validation.

Artificial Neural Network (ANN)

The feedforward neural network used in this study, known as a Multilayer Perceptron, follows a specific type of ANN architecture where information flows unidirectionally from input to output without feedback from the outputs to previous layers. The MLP was configured with two hidden layers, consisting of 30 neurons in the first hidden layer and 20 neurons in the second hidden layer. The logistic sigmoid activation function was utilized in the hidden layers. For training, the Levenberg-Marquardt algorithm was selected, which utilizes backpropagation to iteratively adjust the network weights and minimize the difference between the network predictions and the actual weights. This training process involves calculating weight gradients through error retropropagation and updating the weights using an optimization algorithm such as Stochastic Gradient Descent. The ANN was implemented using MATLAB R2022a. The training was conducted for a maximum of 100 epochs to achieve convergence and enhance the predictive performance of the neural network. The number of neurons in the hidden layers was chosen through a process of adjustment and experimentation.

Results and Discussion

We conducted our study using 889 body images of 79 Nelore cattle in their finishing phase in feedlot. Images from each animal were periodically captured as well as their body weight (BW), independently measured using passive scales. Our primary focus was to assess the effectiveness of extracting Fourier-based features from these images for predicting the animals' BW. To achieve this goal, we compared the Fourier-based features with a comprehensive set of 49 biological features commonly employed in the livestock literature for live weight estimation. These biological features encompassed various measurements from the top-view images, including body

area, major axis, minor axis, eccentricity, circularity, equivalent diameter, solidity, perimeter, and roundness, as well as 20 width (W) and 20 length (L) measurements.

In our initial exploratory analysis, we calculated the correlations between the biological features and the actual weight of the animals. Those features with correlations with p-values smaller than 5% are depicted in Figure 3.4. This figure displays W and L measurements in grayscale, emphasizing the significant regions on the animals' bodies. Out of the 49 biological features examined, 17 displayed statistically significant correlations with the actual weight of the animals ($P < 0.05$). These features include W11, W17, L2, L3, L4, L6, L7, L9, L11, L12, L17, L19, Area, MinorAxisLength, Eccentricity, Solidity, and Perimeter. Notably, in terms of width, the rear and central regions of the animal exhibited stronger correlations with its actual weight. Regarding length, significance was observed in the central region and lateral oscillations of the body. These findings suggest that contours and folds in the animals' body structure may play a crucial role in predicting their weight from such top view images.

The four different models (PLS, RR, LR, and ANN) implemented using the 49 biological features as predictors of weight delivered correlations between the actual and predicted weight ranging from 0.55 to 0.77. Such correlations are relatively lower than those presented by Cominotte et al. (2020), who also utilized images to predict body weight of Nelore cattle in the finishing phase. However, Cominotte et al. (2020), as most computer vision studies available on prediction of cattle weight, used constrained animals for image acquisition. In our case we worked with unrestrained animals and grouped in pens. This imposed a much more challenging condition for image acquisition, given the variable positioning and orientation of the freely moving animals. For example, recent studies, including the work by Hou et al. (2023), emphasize the significant impact of animal movement during data collection on the extraction of these features. During data

collection, animal movements can lead to deviations in the measurement of the shoulder endpoint, resulting in predicted weight values exceeding the actual weight. Conversely, errors in measuring dorsal height can yield predicted weight values below the actual weight.

Nonetheless, our objective is to develop tools that can be applicable to commercial feedlots, so under outdoor conditions and unrestrained groups of animals. This inevitable brings additional challenges, and that is what led us to seek efficient strategies that would allow a deeper insight regarding animal body shapes from their images, specifically through Fourier Transformation. The Fourier Transform applied to the 889 images, each with dimensions of 640x480 pixels, using the formula provided in this paper, generates 889 individual matrices. Each of these matrices is then divided into magnitude and phase components, resulting in two separate matrices, both with dimensions of 640x480. Therefore, a total of 1,778 matrices were generated, each containing 640x480 values. This corresponds to a dataset with $1,778 * 640 * 480$ columns of information. This dataset is rich in details and features extracted from the original images. However, to mitigate dimensionality, we applied Principal Component Analysis (PCA) while retaining 95% of the variability. This led to the extraction of 1,284 components, referred to as Fourier features.

When we employed these 1,284 components as inputs in the four prediction models investigated, we observed an average increase of 0.255 in the correlation between predicted and observed weights (Table 3.2). This suggests that Fourier transform can capture information in the images of the animal bodies that goes beyond the 49 biological features considered. Across all evaluated models, the correlation between model predictions and the actual weight of the animals consistently exceeded 0.85, demonstrating a reasonable predictive capacity. Particularly noteworthy is the combination of the ANN and Fourier Transform, which yielded a highly precise model with 0.98 correlation between predicted and observed weights (Figure 3.5), mean bias of

0.39 kg, and RMSE of 10.89 kg. In comparison with the use of biological features, the ANN exhibited significantly lower prediction quality, with the mean bias approximately 4.3 times lower, RMSE roughly 3.2 times lower, and MAE about 5.3 times lower.

Therefore, despite the challenges of capturing images of animals in motion and under outdoor conditions, our findings demonstrate the ability to achieve comparable or superior results relative to other studies in which images were obtained under more controlled conditions, whether indoors or by confining the animals to specific areas for image collection. For example, when Cominnote et al. (2020) employed CVS to predict body weight and average daily gain in Nelore steers during the growth and finishing phases, they achieved concordance correlation coefficients ranging from 0.81 to 0.95 and RMSEP values ranging from 7.78 to 18.14 (using ANN, PLS, and LASSO). Similarly, Kamchen et al. (2021) used a stereoscopic depth camera to predict body measurements and weight in 260 Nelore heifers, effectively demonstrating the potential to estimate body mass in Nelore heifers using depth-image-derived body volumes, with a correlation coefficient of 0.97 (using ANN), mean absolute error of 8.85 kg, and mean absolute percentage error of 3.13%. Additionally, Gomes et al. (2016) investigated the first generation of Kinect I depth cameras in the context of the relationship between body weight and hot carcass weight (HCW) for 20 Black Angus and 15 Nelore bulls, reporting correlation coefficients ranging from 0.69 to 0.84.

Our findings underscore the potential of employing CVS in real-world feedlot environments. By integrating Fourier analysis into these processes, we significantly enhanced the prediction accuracy of live weights of unrestrained cattle. This may offer a viable alternative for monitoring feedlot animals' weight compared to passive weighing scales located within pens, which suffers from many problems such as variation on body weight given water and feed content on digestive tract, and the need for periodic cleaning and calibration of scales. These limitations

were brought to light in a study conducted by Kelly et al. (2019), which showed significant fluctuations in the body weights of growing cattle over a mere two-day period, with variations spanning from -18 to 22 kg. In this context, the adoption of CVS emerges as a promising strategy which not only offers the possibility of systematic body weight prediction with good accuracy, but also allows monitoring of various aspects of animal behavior, such as feeding and drinking, as discussed in Lind et al. (2005), Kashiha et al. (2013), Bresolin et al. (2022), among others.

Furthermore, the approach can be easily extended to other beef cattle breeds, although the models may need to be retrained using additional data from such breeds. By collecting top-down images of animals from different breeds and subsequently retraining the model, we can empower our system to extract Fourier features from previously unobserved breeds. This strategic development enhances the system's potential to facilitate the identification of specific breeds and, consequently, enables automated weight prediction for a broader spectrum of cattle.

Monitoring body weight of cattle in feedlots can be extremely useful for helping better feeding management, as also for early detection of sick animals or metabolic issue. However, for helping determining the best harvest time of individual animals or groups of animals (i.e. pens), it would be interesting also to monitor muscle development and fat deposition. In this context, it is indeed our interest of future research on using animal images to assess body composition, as in Fernandes et al. (2020). A potential way to achieve this using our approach would be by exploring the magnitude spectrum profile, which represents the mean of all sinusoidal waves extracted using Fourier transformation. Figure 3.6A illustrates an example of this curve, with its primary peak closely associated with total body mass. Also, less prominent waves indicate additional peaks at the tails of the curve. These patterns might be potentially connected with the animal body contours, which are related to muscle development and fat deposition.

Hence, an idea would be to quantify and examine the mean spectrum profile of cattle by calculating the area under the most prominent curve (total area) and the less prominent peaks (contour) of the magnitude spectrum profile. The relationship between the size of the central 'peak' and the size of the 'tails' of the curve can serve as a metric to evaluate changes in body composition. A smaller 'peak' in relation to the 'tails' may suggest a reduction in body fat and an increase in muscle density.

To emphasize the practical application of this analysis, consider a set of six theoretical ellipsoidal shapes (Figure 3.6B), ranging from a 'flatter' ellipse to a more 'elongated' one, while keeping the size of the major axis constant. This illustration aims to demonstrate how analyzing the mean magnitude spectrum of sinusoidal waves could potentially help assessing the animal's body development. Analyzing the mean magnitude spectrum of these shapes (Figure 3.6C), we notice that these curves effectively reflect changes in the object's shape. For instance, in the first shape, with a 'flatter' ellipse, the peak is less prominent, and the tails of the curve are less pronounced. As we progress to more 'elongated' shapes, the central 'peak' decreases in size, while the tail regions of the curve become more pronounced, thereby capturing the details of shape variation.

Conclusions

In this study, we introduced an innovative method for extracting body weight features based on the Fourier transform, to be used as input in prediction models for body weight assessment of feedlot cattle. Our approach yields competitive results compared to traditional methods of measuring image features. Moreover, our findings indicate that the combination of Fourier transform with Artificial Neural Networks (ANNs) holds great promise for accurately

predicting live weight of cattle in feedlot conditions, which can significantly benefit feedlots by enabling daily weight monitoring and facilitating improved management practices. Our work underscores the flexibility and promise of computer vision applications in livestock management, paving the way for increased accuracy and efficiency in the livestock industry.

Table 3.1 Descriptive statistics of body weight of beef cattle in feedlot.

| Statistics | Weight Data |
|------------------------------|-------------|
| mean (kg) | 507.94 |
| standard deviation (kg) | 54.58 |
| minimum (kg) | 390.99 |
| maximum (kg) | 609.03 |
| coefficient of variation (%) | 10.74 |
| images (n) | 889 |
| animals (n) | 79 |

Table 3.2 Performance of weight prediction models for beef cattle using different image features.

| Features | Mean bias | RMSEP ^a | RMSEP% ^b | MAE ^c | Pred-obs Corr ^d (95% CI ^e) |
|---|-----------|--------------------|---------------------|------------------|---|
| Artificial Neural Network (ANN) | | | | | |
| Biological | 4.69 | 34.88 | 6.87 | 27.87 | 0.77 (0.75 0.80) |
| Fourier | 0.39 | 10.89 | 2.14 | 5.22 | 0.98 (0.97 0.99) |
| Ridge Regression (RR) | | | | | |
| Biological | -0.05 | 43.74 | 8.61 | 36.18 | 0.60 (0.55 0.64) |
| Fourier | -0.05 | 26.14 | 5.15 | 20.98 | 0.88 (0.86 0.89) |
| Lasso Regression (LR) | | | | | |
| Biological | -0.01 | 43.84 | 8.63 | 36.23 | 0.60 (0.55 0.64) |
| Fourier | -0.03 | 28.87 | 5.68 | 22.86 | 0.86 (0.84 0.87) |
| Partial Least Square Regression (PLSR) | | | | | |
| Biological | -5.61 | 44.99 | 8.86 | 35.24 | 0.58 (0.54 0.62) |
| Fourier | -1.38 | 28.89 | 5.69 | 23.01 | 0.85 (0.83 0.87) |

a: Root Mean Square Error of Prediction; b: Root Mean Square Error of Prediction Percentage; c: Mean Absolute Error; d: correlation between the predicted weight and the observed weight; e: 95% confidence interval.

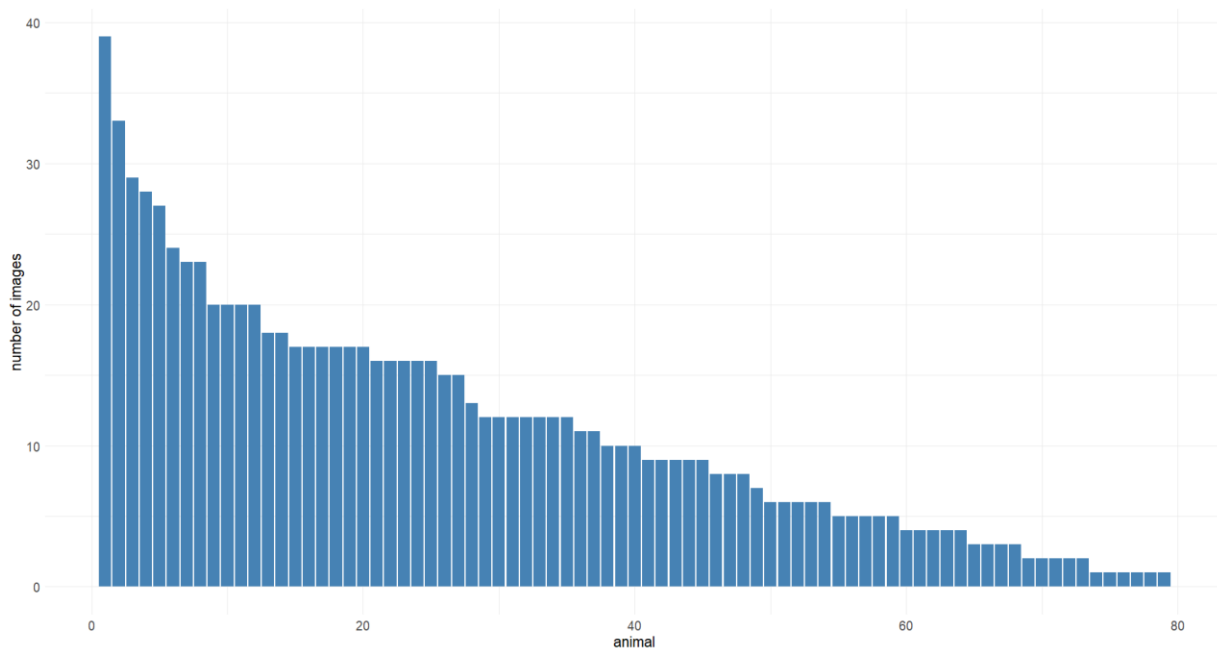


Figure 3.1 Frequency distribution of images for each of the 79 animals in our experiment.

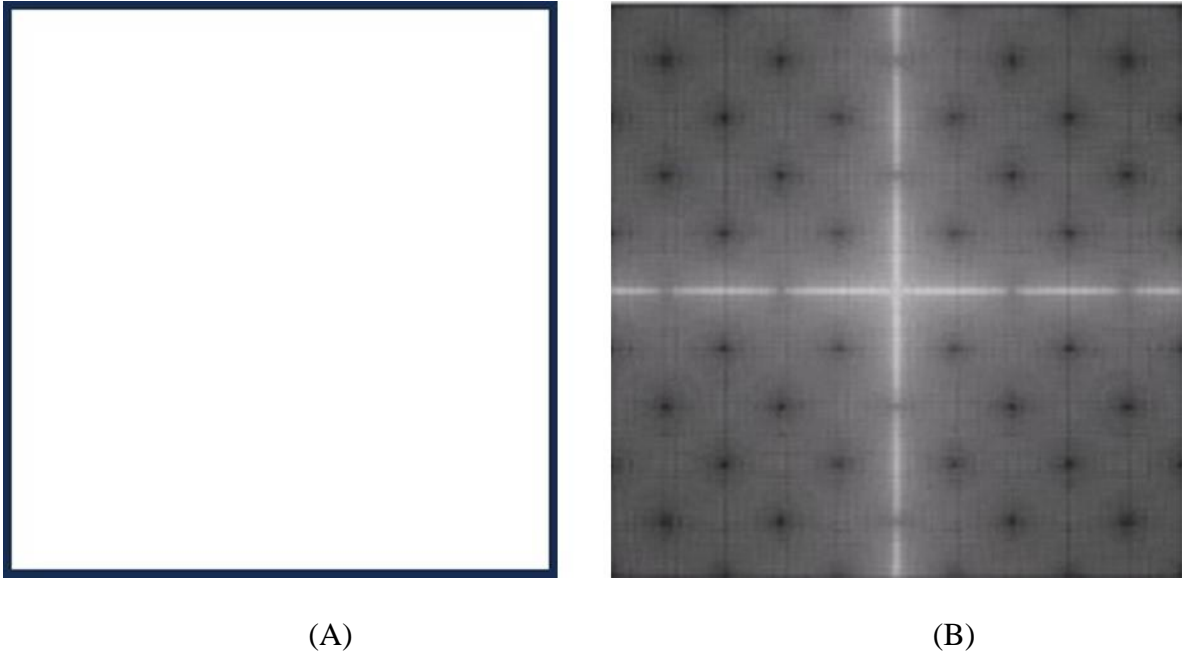


Figure 3.2 (A) Binary image of a square. (B) Magnitude spectrum of the Fourier Transform for the binary image of the square.

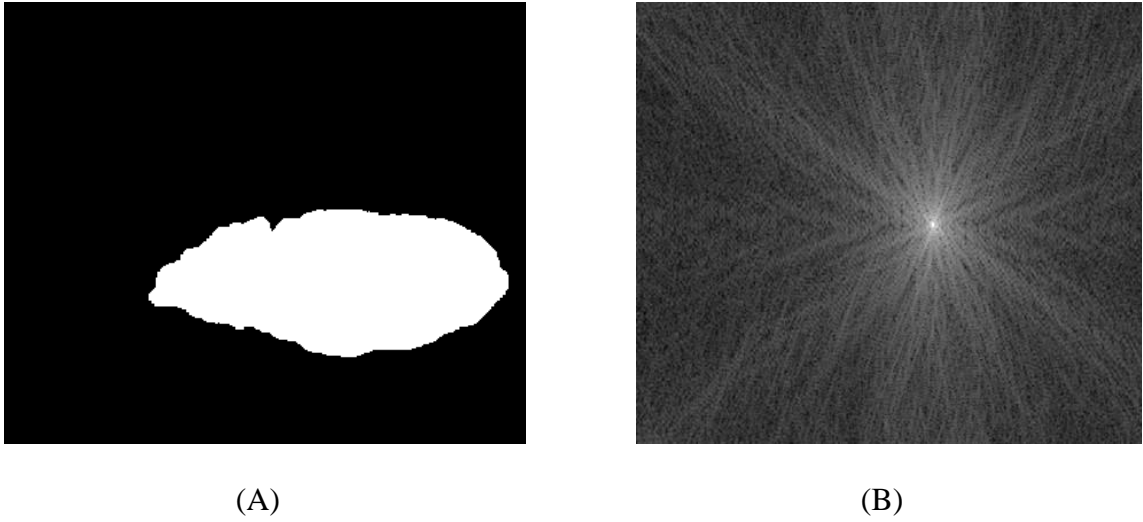
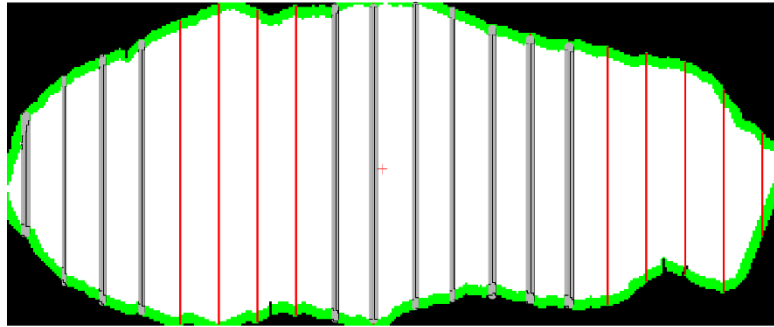
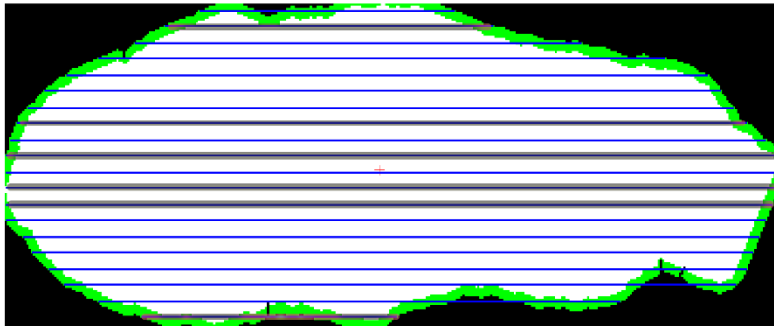


Figure 3.3 (A) Binary image of the animal's body from Top View. (B) Magnitude spectrum of the Fourier transform for the binary image of the animal's body.



A



B

Figure 3.4 Top-view image of a beef cattle with 20 width and length measurements used as predictors in a multiple linear regression model with real weight as the response variable. Gray highlighting indicates the significant Pearson correlation between the feature and the actual weight ($P < 0.05$). (A) Width measurements; (B) Length measurements.

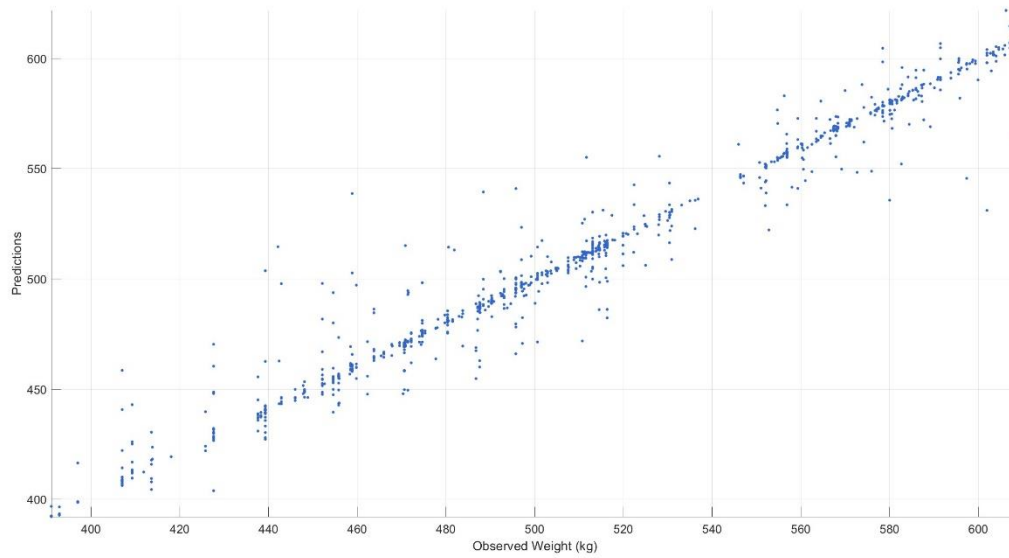
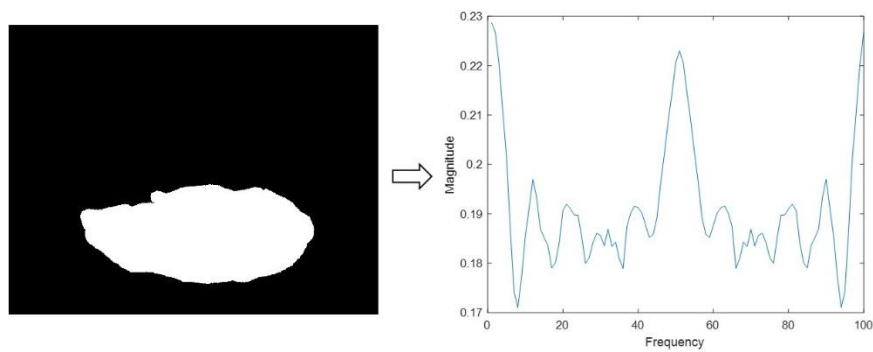


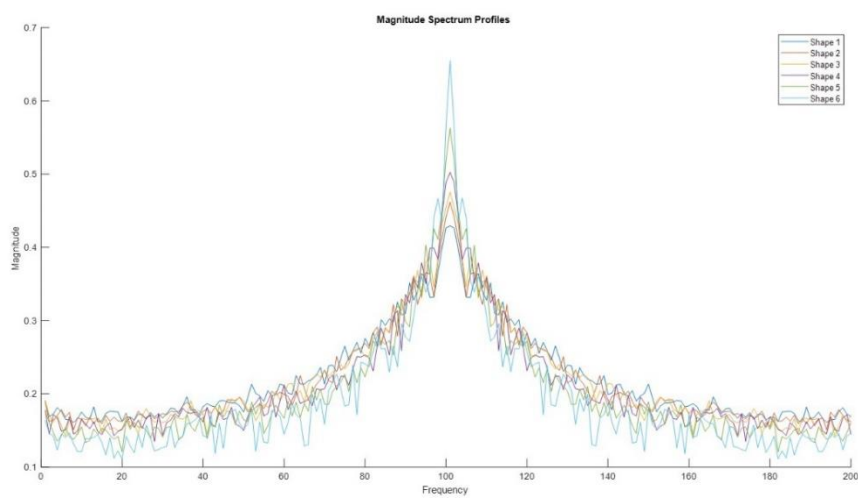
Figure 3.5 Correlation between observed and predicted body weight using Fourier features as predictors in the ANN model.



(A)



(B)



(C)

Figure 3.6 (A) Example of a magnitude spectrum profile for an animal body image. (B) Six theoretical ellipsoidal shapes, ranging from a 'flatter' ellipse to a more 'elongated' one, while keeping the size of the major axis constant. (C) Magnitude spectrum profiles for the six ellipsoidal shapes.

References

- Bresolin T, R. Ferreira, F. Reyes, J. Van Os, and J. R. R. Dorea. 2023. Assessing optimal frequency for image acquisition in computer vision systems developed to monitor feeding behavior of group-housed Holstein heifers. *J. Dairy Sci.* 106:664-675. doi: 10.3168/jds.2022-22138.
- Cominotte, A., A. F. A. Fernandes, J. R. R. Dorea, G. J. M. Rosa, M. M. Ladeira, E. H. C. B. Van Cleef, G. L. Pereira, W. A. Baldassini, and O. R. Machado Neto. 2020. Automated computer vision system to predict body weight and average daily gain in beef cattle during growing and finishing g phases. *Livest. Sci.* 232:103904-103914. doi: 10.1016/j.livsci.2019.103904.
- Cooke, R. F., and E. Bill. 2014. Kunkle Interdisciplinary Beef Symposium: Temperament and acclimation to human handling influence growth, health, and reproductive responses in *Bos taurus* and *Bos indicus* cattle. *J. Anim. Sci.* 92:5325–5333. doi:10.2527/jas.2014-8017.
- Cooley, J. W., and J. W. Tukey. 1965. An Algorithm for the machine calculation of complex Fourier series. *Math. Comput.* 19:297-301.
- Dohmen, R., C. Catalb, and Q. Liua. 2022. Computer vision-based weight estimation of livestock: a systematic literature review. *N. Z. J. Agric. Res.* 65:227–247. doi:10.1080/00288233.2021.1876107.
- Dorea, J. R. R., and G. J. M. Rosa. 2022. Computer vision systems to advance high-throughput phenotyping in livestock. *Proc. 12th World Congress on Genetics Applied to Livestock Production (WCGALP)*. 138: 602 – 605.
- Faraggi, M. 2019. Time series features extraction using Fourier and Wavelet transforms on ECG data. *Octo Blog*. Retrieved January 6, 2023, from [<https://blog.octo.com/time-series-features-extraction-using-fourier-and-wavelet-transforms-on-ecg-data/>].
- Fernandes, A. F. A., J. R. R. Dorea, B. D. Valente, R. Fitzgerald, W. Herring, and G. J. M. Rosa. 2019. A novel automated system to acquire biometric and morphological measurements and predict body weight of pigs via 3D computer vision. *J. Anim. Sci.* 97:496–508. doi: 10.1093/jas/sky418.
- Fernandes, A. F. A., J. R. R. Dórea, R. Fitzgerald, W. Herring, and G. J. M. Rosa. 2019. Comparison of data analytics strategies in computer vision systems to predict pig body composition traits from 3D images. *J. Anim. Sci.* 98:8, skaa250. doi: 10.1093/jas/skaa250.
- Fernandes, A. F. A., J. R. R. Dorea, and G. J. M. Rosa. 2020. Image analysis and computer vision applications in animal sciences: An overview. *Front. Vet. Sci.* 7:551269. doi:10.3389/fvets.2020.551269.

- Gomes, R. A., G. R. Monteiro Assis, G. J. F. Busato, M. M. Ladeira, and M. L. Chizzotti. 2016. Estimating body weight and body composition of beef cattle through digital image analysis. *J. Anim. Sci.* 94:5414-5422. doi:10.2527/jas.2016-0797.
- Guesmi, H., H. Trichili, A. M. Alimi, and B. Solaiman. 2015. Fingerprint verification system based on curvelet transform and possibility theory. *Multimed Tools Appl.* 74:3253-3272. doi:10.1007/s11042-013-1785-1.
- Haskell, M. J., G. Simm, and S. P. Turner. 2014. Genetic selection for temperament traits in dairy and beef cattle. *Front. Genet.* 5:368. doi:10.3389/fgene.2014.00368.
- Hassen, H., and K. Maher. 2014. Distributed Fast Fourier Transform (DFFT) on MapReduce Model for Arabic Handwriting Feature Extraction Technique via Cloud Computing Technologies. *Proc. Conference on Parallel and Distributed Processing Techniques.*
- Heinrichs A.J., G.W. Rogers, and J.B. Cooper. 1992. Predicting body weight and wither height in Holstein heifers using body measurements. *J. Dairy Sci.* 75:3576-3581. doi:10.3168/jds.S0022-0302(92)78134-X.
- Hoerl, A.E., and R. W. Kennard. 1970a. Ridge Regression: Biased estimation for nonorthogonal problems. *Technometric.* 12:55-67. doi:10.1080/00401706.1970.10488634.
- Hoerl, A.E., and R. W. Kennard. 1970b. Ridge Regression: Applications to nonorthogonal problems. *Technometrics.* 12:69-82. doi:10.1080/00401706.1970.10488635.
- Hou, Z., L. Huang, Q. Zhang, and Y. Miao. 2023. Body weight estimation of beef cattle with 3D deep learning model: PointNet++. *Comput. Electron. Agric.* 213:108184. doi:10.1016/j.compag.2023.108184.
- Kamchen, S. G., E. F. dos Santos, L. B. Lopes, L. G. Vendrusculo, and I. C. Condotta. 2021. Application of depth sensor to estimate body mass and morphometric assessment in Nellore heifers. *Livest. Sci.* 245:104442. doi:10.1016/j.livsci.2021.104442.
- Kashiha M. A., C. Bahr, S. A. Haredasht, S. Ott, C. P. H. Moons, T. A. Niewold, and O. F. Odberg, and D. Berckmans. 2013. The automatic monitoring of pigs water use by cameras. *Comput. Electron. Agric.* 90:164–9. doi:10.1016/j.compag.2012.09.015.
- Kelly, D. N., C. Murphy, R. D. Sleator, M. M. Judge, S. B. Conroy, and D. P. Berry. 2019. Feed efficiency and carcass metrics in growing cattle. *J. Anim. Sci.* 97:4405–4417. doi:10.1093/jas/skz316.
- Lees, A. M., H. E. Salvin, I. G. Colditz, and C. Lee. 2020. The influence of temperament on body temperature response to handling in Angus cattle. *Animals.* 10:172. doi:10.3390/ani10010172.

- Lind N. M., M. Vinther, R. P. Hemmingsen, and A. K. Hansen. 2005. Validation of a digital video tracking system for recording pig locomotor behaviour. *J. Neurosci. Methods* 143:123–32. doi: 10.1016/j.jneumeth.2004.09.019
- MATLAB R2022a. (Version 10.2.1). MathWorks. Retrieved from <https://www.mathworks.com/products/matlab.html>.
- Mironovova, M., and J. Bíla. 2015. Fast Fourier transform for feature extraction and neural network for classification of electrocardiogram signals. *Proc. Conference on Future Generation Communication Technologies (FGCT)*. 4:1-6. doi: 10.1109/FGCT.2015.7300244.
- Mohammady, S. 2021. *Wavelet Theory*. IntechOpen. doi: 10.5772/intechopen.87895.
- Oliveira, D. A. B., L. G. R. Pereira, T. Bresolin, R. E. P. Ferreira, and J. R. R. Dorea. 2021. A review of deep learning algorithms for computer vision systems in livestock. *Livest. Sci.* 253:104700. doi: 10.1016/j.livsci.2021.104700.
- Ricker, N. H. 1940. The form and nature of seismic waves and the structure of seismograms. *Geophysics*. 5:348-366.
- Rudenko, O. G., Y. Megel, O. Bezsonov, and A. Rybalka. 2020. Cattle breed identification and live weight evaluation on the basis of machine learning and computer vision. *Proc. International Workshop on Computer Modeling and Intelligent Systems*. 2608:70.
- Schofield, C. P., J. A. Marchant, R. P. White, N. Brandl, and M. Wilson. 1999. Monitoring pig growth using a prototype imaging system. *J. Agric. Eng. Res.* 72:205–210. doi:10.1006/jaer.1998.0365.
- Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc.* 58:267-288.
- Ulutas Z, M. Saatci, and A. Ozluturk. 2002. Prediction of body weights from body measurements in East Anatolian Red calves. *Indian J. Anim. Sci.* 72:878-881.
- Wang, Z., S. Shadpour, E. Chan, V. Rotondo, K. M. Wood, and D. Tulpan. 2021. ASAS-NANP SYMPOSIUM: Applications of machine learning for livestock body weight prediction from digital images. *J. Anim. Sci.* 99:1–15. doi: 10.1093/jas/skab022.

CHAPTER FOUR: EXPLORING DEEP LEARNING FOR FACE-BASED IDENTIFICATION
IN BEEF CATTLE CONSIDERING TWO DIFFERENT LOCATIONS OF THE
PRODUCTION SYSTEM.

Abstract

The global beef industry stands as a vital component of the world's food system, supporting food security and economic development. Efficient management of animal identification is crucial for the ongoing success of the industry. This study explored the potential of Convolutional Neural Networks (CNN) in the identification and differentiation of animals within and across two distinct locations in the cattle industry using facial images. The research involved a diverse group of 25 animals, including those with easily distinguishable features and challenging cases, video recorded in a feedlot exit (Location 1) and a slaughterhouse entrance (Location 2). The results revealed that when CNN-based classification models are trained and tested using images from the same location, they achieve notable accuracy levels of 0.853 (*ResNet50*) and 0.752 (*Xception*), highlighting their effectiveness within a single location. However, challenges arised when these models are subjected to cross-location testing, resulting in decreased accuracy rates of 0.277 (*ResNet50*) and 0.207 (*Xception*). These challenges predominantly stem from the scenario of training in one location and testing in another, highlighting the need for consistent pre-processing and image alignment, encompassing factors such as camera angle, lighting, and background. While the relatively small dataset poses constraints, this preliminary investigation had not only illuminated our approach's limitations but also brought to light its strengths. It has guided our understanding of specific areas needing improvement. As a result, this study served as a pivotal steppingstone, providing a clear

direction for future enhancements in image processing and the exploration of alternative model architectures for image-based animal identification and tracking.

Introduction

The beef industry plays a prominent role in the global context, serving as one of the cornerstones of the world's food system. Its significance extends beyond providing a crucial source of animal protein; it is a fundamental pillar of the global economy, contributing significantly to the livelihoods of millions of people worldwide (OECD/FAO, 2023). The growing demand for beef is intricately linked to population growth and the relentless pursuit of sources of essential nutrients. Furthermore, the unique ability of cattle to convert inedible biomass into high-value biological protein makes this industry essential in addressing the challenges of food security and sustainable food supply worldwide (Baber et al., 2018).

To ensure the continuous success of this sector, effective management of animal identification is imperative. This is because, before reaching the shelves of supermarkets and butcher shops, each animal passes through a complex supply chain that may involve farmers, breeders, processing units, and internal transfers within facilities. In this context, crucial issues emerge, such as ensuring precise animal origin, compliance with regulations in both domestic and export markets, as well as optimizing operational efficiency. Counting errors in the context of herd inventory, for example, can result not only in significant financial losses but also in management inefficiencies when dealing with discrepancies.

Currently, significant investments have been made in the livestock industry to implement Technology in Traceability Management (TTM) programs to enhance animal tracking. These technologies offer various tracking approaches, such as the use of GPS, Bluetooth or Wi-Fi

networks, and Radio-Frequency Identification (RFID) technologies for animals (Huhtala, 2007). Among these techniques, RFID stands out as one of the most promising approaches for beef cattle. While the industry has made significant investments in modernizing animal identification using RFID devices, these technologies can be costly and pose challenges in terms of maintenance and management. Additionally, they are susceptible to being lost, removed, or damaged (Gaber et al., 2016).

In this context, computer vision methods have emerged as a viable and continually evolving alternative to address these challenges in the livestock industry. This application for animal identification, with roots dating back to the 1990s, has gained substantial momentum over the decades. Pioneering studies, such as those conducted by McFarlane and Schofield (1995) and Tillett et al. (1997), explored pig identification during those decades, while Sergeant et al. (1998) applied computer vision techniques to track poultry. In recent years, this field has grown significantly, with various applications in beef cattle (Bezen et al., 2020; Kumar et al., 2018; Qiao et al., 2019) and swine (Hansen et al., 2018; Marsot et al., 2020; Zheng et al., 2015), especially for animal identification in feedlot conditions. For instance, Li et al. (2022) demonstrated the potential of using deep learning techniques to identify individual cattle based on muzzle images.

These approaches aimed to leverage unique individual characteristics, such as body patterns, facial features, muzzles, or iris patterns. Notably, none of these studies had specifically addressed the problem of identifying the same animal through facial images captured at various locations along its journey within the meat production chain. In this emerging and challenging field, we proposed an initial approach to gain insights into the nuances, challenges, and specific characteristics related to cattle's face identification using Convolutional Neural Networks (CNN).

Hence, our objective in this study was to assess the system's ability to identify and distinguish animals within the same location and across two different locations within the cattle production chain using facial images. To achieve this, we conducted an experiment in which we selected a feedlot to represent Location 1 and a slaughterhouse for Location 2. Initially, we recorded videos of three truckloads at Location 1, totaling 100 animals. These same trucks were documented upon their arrival at Location 2. Additionally, we captured images from various other trucks originating from different cattle providers at Location 2. Our focus was on a specific group of 25 animals. These 25 animals were visually identified as appearing at both locations using images of their faces obtained through YOLO, by watching the original videos and checking the truck license plates. Some of these animals had distinctive characteristics, such as spots, specific colors, or notable physical features, including the presence of horns and unique markings on their bodies, while others belonged to homogeneous breeds, sharing similar traits, such as color. Hence, our specific objective was to assess the effectiveness of two different CNN architectures for individual face recognition in beef cattle. Hence, our specific objective was to assess the effectiveness of two different CNN architectures for individual face recognition in beef cattle.

Materials and Methods

Data Acquisition

In this study, we conducted a pilot experiment at an important beef processing company facility in Brazil. We strategically selected two points for our research: Location 1, where animals were loaded from the feedlot to the transportation trucks, and Location 2, where animals were unloaded from the trucks to the slaughterhouse. During one full week of image collection, we were able to video record three truck loads at the feedlot (Location 1), comprising 100 animals. The

same three truck loads were recorded at arrival at the slaughterhouse (Location 2). In Location 2, in addition to the three trucks coming from this specific feedlot, we also captured images from a number of other trucks coming from other cattle providers; these trucks however were not used for the purpose of this study, which was focused on animals appearing on both locations. On both locations we positioned RGB cameras in strategic locations about 3 meters high to capture the frontal part of the animals. At the feedlot facility, the camera was located on the corridor between the corral and the ramp leading to the truck load. At the slaughterhouse, the camera was located on the tunnel leading to the rest area, about 5 meters from the truck door through which animals are unloaded. Examples of images captured in each location are illustrated in Figure 4.1.

At Location 1, we used the Dome Intelbras VHC 1120 D camera, connected to a computer on the same network with internet access, to record videos every 10 minutes throughout the day. For Location 2, we employed the Hikvision DS-2CD2345FWD-I camera, connected to a computer without internet access, and videos were triggered only when trucks left the confinement area. We manually selected videos for analysis, focusing on those containing animals in the scene.

After obtaining the videos, we proceeded with the training and application of a YOLO (You Only Look Once) model to identify and crop the face of each animal in both locations. This resulted in a total of 5,368 RGB frames with an average resolution of 270×230 pixels. Figure 4.2 illustrates the process of image detection and cropping using the YOLO network, and it also provides some examples of the cropped images from both locations.

To select animals that appeared at both locations, we conducted a manual identification process to establish correspondences. This involved three distinct data verification procedures to ensure the integrity and accuracy of the established correspondences. Initially, we utilized cropped images provided by the YOLO model, followed by a meticulous analysis of the original videos.

Additionally, we conducted an additional verification step, which included inspecting the license plates of the trucks transporting the animals to confirm that the vehicles leaving the feedlot were indeed the same ones arriving at the slaughterhouse. This comprehensive process identified 25 correspondences, which were the primary focus of this study, resulting in a total of 215 images. Among these, 80 originated from the feedlot, while 135 were from the slaughterhouse, as the YOLO system generated multiple images per animal.

Data Analysis

In this study, we explored animal identification in a closet-set context based on images of their faces using two well-known Convolutional Neural Networks (CNN) architectures: *ResNet50* (RN50) and *Xception* (XPT). We annotated 215 images for 25 labels, each representing the identity of a specific animal. We expect that these classifiers can learn to recognize the individual characteristics of each animal during training and apply the generated feature mapping to infer the animals' identifications in yet-to-be-seen images.

We employed two strategies in this study. The first strategy involves training the networks with images from the same location (Location 2) and then apply them to unseen images during training to evaluate their identification ability within the same training location. The second is to train the networks with images from one location (Location 2) and testing them on images from another location (Location 1) to assess their recognition capability.

To evaluate the performance of the networks, we conducted a five-fold cross-validation process, randomly dividing the images generated at location 2 into training (80%) and testing (20%) sets. Also, for each training set a second testing set was created by randomly drawing 80% of the images generated at the feedlot (Location 1). During training, we applied a technique called

data augmentation, creating ten-fold replications of each image with slight variations, such as random rotation, brightness, and zoom. This process was performed using the *ImageDataGenerator* functionality from Keras, with the following parameters: `rotation_range = 20`, `zoom_range = 0.13`, `brightness_range = (0.8, 1.5)`, `horizontal_flip = True`, `vertical_flip = False`, `fill_mode = 'nearest'`. This resulted in an average of 2300 images in the training set, while the test images remained unaltered. Further details of the data analysis methods are described later in this paper.

Object Detection and YOLO

Object detection is a computer vision technique employed to identify and locate objects within images or videos, facilitating precise counting, tracking, and labeling within a scene. Our research focuses on single-stage deep learning-based algorithms, exemplified by YOLO (You Only Look Once), which utilize regression techniques to predict object bounding boxes and label probabilities in a single network pass. The term 'You Only Look Once' refers to an object detection method that prioritizes efficiency by performing all detection steps in a single pass through the neural network. In other words, instead of examining the image multiple times at different scales and resolutions, YOLO analyzes the image just once, resulting in faster and more effective object detection.

Due to its efficiency, YOLO has gained popularity in the field of animal detection and tracking. In a review conducted by Borges Oliveira et al. (2021), various applications were highlighted, with a primary focus on animal detection, especially in swine (Cowton et al., 2019; Lee et al., 2019; Psota et al., 2019; Seo et al., 2020), but also in dairy cattle (Kang et al., 2020;

Cernek et al., 2020). Shao et al. (2023) and Islam et al. (2023) have recently contributed to cattle identification and tracking.

Since its inception in 2015, YOLO has undergone several iterations, with a focus on optimizing its efficiency to provide real-time classification performance, even with limited computational resources. The latest version, YOLO-v8, was released in January 2023 and is the one used in this work through Python. To train this network, we meticulously annotated 573 images, and conducting rigorous testing on a separate set of 272 images. The annotation process involved manual delineation of cattle heads in both the training and testing datasets. To ensure uniformity in image dimensions, we resized all input images to 640 x 640 pixels. Subsequently, we initiated the training phase, running the algorithm for a total of 50 epochs. To assess the model's performance, we analyzed the F1-confidence curve, precision-recall curves, and mAP (mean Average Precision) for both the training and test datasets.

Animal Identification via Convolutional Neural Networks (CNN)

The outputs from the YOLO model were then used as the input of two different CNN architectures: *ResNet50* (He et al., 2016) and *Xception* (Chollet, 2017). A ResNet (Residual Neural Network) architecture has a block with two layers, in which the input that passes through the first layer is added to the original input, and then goes through the second layer. This is known as a "skip connection" or "shortcut connection." This connection allows the network to learn the difference between the original input and the transformations performed by the first layer before proceeding to the second layer. This approach makes training easier and more effective, enabling the construction of deeper networks without gradient vanishing issues. *ResNet50*, as the name suggests, is a deep neural network with 50 layers and excels in tasks related to image classification

(He et al., 2016). *Xception* is an advanced extension of the Inception architecture that introduces an innovative approach using separable convolutions to enhance feature representation in images. In contrast to traditional CNN where sequential convolutional layers can lead to gradient-related issues, this network maintains continuous connections between its layers. With 36 convolutional layers organized into 14 modules, the architecture employs 11 depth-wise convolutions. This results in a richer and more detailed representation of image features. Therefore, *Xception* is capable of consolidating information from multiple input channels to form a cohesive image, thereby improving its classification capability (Chollet et al., 2017).

For the two CNN architectures used as the backbones of our classifiers, the last fully connected (FC) layer was removed and all the remaining layers were initialized using weights obtained from the respective networks trained using the ImageNet dataset (Deng et al., 2009). This is an open image dataset that contains more than 1 million examples of diverse environments and objects such as wild and farm animals, humans, and cars. Our objective was to leverage the weights previously optimized in a different task domain using a much larger dataset to help our task-specific network to learn generic features, such as textures, edges, corners, and shapes. This technique is commonly referred to as Transfer Learning (Weiss et al., 2016). Given the small size of our dataset, both architectures were expanded with Global Average Pooling (GAP) instead of flattening the last layer, generating 2048 features in both *ResNet50* and *Xception* architectures, hereinafter referred to as RN50 and XPT, respectively. The generated features were connected to an FC layer of size 256 with the Rectified Linear Unit (ReLU) being used as an activation function followed by an output layer of size equal 25 (number of classes) and SoftMax activation function.

For each CNN, the training process was divided into two steps, where during the first training stage only the weights of FC layers were updated for 50 epochs using a learning rate =

0.001. After this initial training stage, the remaining layers were unfrozen and the weights for the whole CNN architecture were optimized for a further 200 epochs using learning rate = 1×10^{-5} . During all stages the Adam optimization algorithm (Kingma and Ba, 2014) was used to minimize the categorical cross-entropy function, assuming a batch size of 16. Figure 4.3 illustrates the overall CNN architecture in our study. Images traverse multiple layers, including convolution, pooling, and dense layers, before reaching the output layer, which generates probabilities associated with different animal identifications. The final class is determined based on the highest calculated probability.

Results and Discussion

Out of the 25 selected animals, 18 exhibited easily distinguishable characteristics, forming a strong foundation for evaluating the classification models' ability to identify individuals with notable visual features. In addition to some distinct animals, there were also some Nellore and Nellore crossbred animals that exhibited significant similarities, thus posing additional challenges. (Figure 4.4).

The selection of 25 animals played a crucial role in this study. A diverse group of animals was chosen, each possessing unique features such as spots, specific colorations, or distinctive physical attributes like horns. Our objective was to provide the classification models with a mix of easily distinguishable cases and more challenging instances, reflecting a broad spectrum of characteristics encountered in real-world animal identification scenarios, encompassing situations where distinctiveness arises from breed diversity as well as scenarios characterized by breed homogeneity.

The YOLO model successfully recognized and cropped the faces of the animals with high accuracy, presenting an F1-score of 0.94 and an mAP (mean Average Precision) of 0.97 both the validation and test datasets for animal face recognition (Figure 4.5 A, B, and C). It's worth noting that we intentionally set a lower IoU threshold of 0.4. This deliberate choice, which allows for a less stringent matching of detected objects, resulted in an increased number of detected instances of animal faces within the video, as illustrated in the frequency distribution graph in Figure 4.6. At the slaughterhouse unit (Location 2), we observed an average of approximately 5 repetitions per animal, with a standard deviation of 3, while in the feedlot (Location 1), the average was about 3 repetitions per animal, with a standard deviation of 1. These repetitions provided diverse angles of the animals, as illustrated in Figure 4.7.

The results of the first training strategy, which involved using images from Location 2 to test at Location 1, revealed an average accuracy of 0.277 (± 0.066) for RN50 and 0.207 (± 0.046) XPT as shown in Figure 4.8. In Figure 4.9, we present heatmaps for the confusion matrices obtained with both models and testing locations. Deeper shades of red in the heatmap highlight animals that were accurately identified, as discussed in the subsequent section.

In the scenario where the CNN were trained with images from Location 1 and tested at Location 2, animals 17, 20, and 21 achieved an accuracy exceeding 0.8 (Figure 4.10). These animals exhibited easily identifiable attributes, such as spots and horns. In contrast, XPT performed less satisfactorily in classifying animals 17 and 20 but demonstrated superior performance with other animals, including numbers 1, 10, and 21, as shown in Figure 4.11. Notably, XPT successfully identified an animal with distinctive horns and another with facial spots, even under challenging conditions with varying lighting at both locations.

Each network had its unique strengths and weaknesses in identifying specific image characteristics. RN50 excelled in extracting differentiating details in the image, while XPT seemed to work better in capturing general features and handling heterogeneous lighting and positioning. Images where both networks failed to achieve satisfactory matches often involved cases with fewer than four repetitions per animal or included repetitions that diverted focus from the face, such as the presence of parts of other animals or significantly different angles.

This leads us to consider various issues related to image capture angles at both locations. Image alignment is necessary when processing these images, even when taken from different angles or with distinct resolutions and lighting. This alignment process is known as image registration, involving aligning two or more images based on their visual appearances (Fu et al., 2019). Registration may be required when analyzing pairs of images taken from different viewpoints, at different times, or using different sensors/modalities (Hill et al., 2001; Zitova and Flusser, 2003). There are various strategies for performing registration, such as DIC (digital image correlation) and methods involving Fourier analysis, and keypoint identification using algorithms like SIFT (Scale-Invariant Feature Transform) and SURF (Speeded-Up Robust Features). Recent studies propose image registration using network models (Fu et al., 2019; Haskins et al., 2020).

Another limitation of this study is the reduced data size used for training the CNN, which prevents the networks from learning sufficient nuances that can generalize well in new images. A potentially better approach in these scenarios could be the use of one-shot or few-shot learning techniques such as Siamese networks (Bromley, J et al. 1993). The idea behind Siamese networks is that they learn a similarity function from pairs of images, without classifying images themselves. Instead, they determine if two input images belong to the same class (e.g. same animal identification). Once the Siamese network is trained, its discriminatory power can be generalized

to classify not only new images of the classes used during training but also classes not seen in the training phase. A recent work in this context addressing the challenge of identification based on small datasets was developed by Figueroa-Mata et al. (2020).

When we trained the models at Location 2 and tested them at the same location, we noted that the accuracy for both RN50 and XPT consistently remained above 0.80 in all five iterations. This result is particularly promising, considering the relatively small size of our dataset. It suggests that animal identification within the same location can be achieved with significant accuracy, which is a relevant result for future applications of non-invasive identification methods in the meat industry.

These results also indicate that if we can improve the homogeneity in image capture angles and perform additional work to remove unwanted backgrounds, we can further enhance Strategy 1, allowing for effective animal tracking across two different locations. This demonstrates the optimization potential of our approach for future advancements in animal identification.

Conclusions

In this study, our primary objective was to assess the effectiveness of two widely used Convolutional Neural Networks (CNN) for image classification in the field of animal identification and differentiation within the livestock industry. Our findings demonstrate the promising potential of these CNN for animal identification within the same location and, even in the face of inherent complexities associated with real-world identification scenarios in different locations, we have garnered valuable insights on improving the process. We observed that optimizing image capture angles, enhancing image quality, and exploring more suitable neural networks for smaller datasets could pave the way for even more promising results in the future. Future work to extend this study

should involve a larger data set, not only in terms of number of animals but also locations with more diversity of environmental conditions.

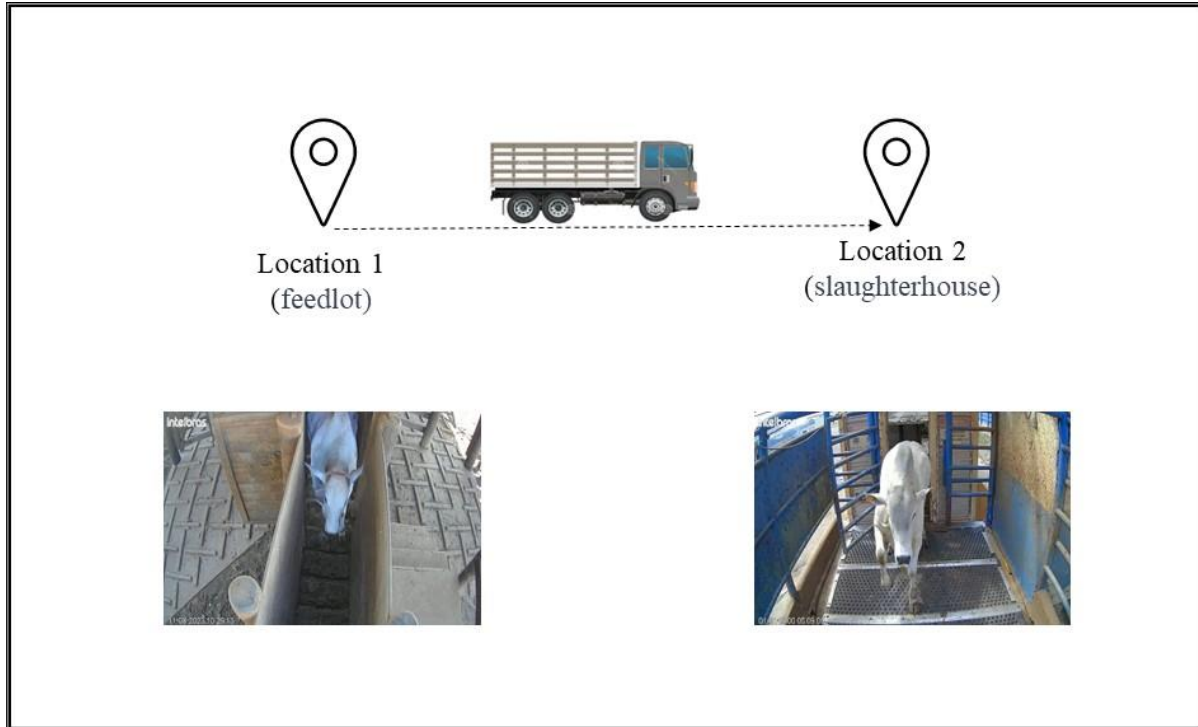


Figure 4.1 Video collection example at two distinct points in the industry: (A) Location 1, the exit of animals from the feedlot. (B) Location 2, the entrance of animals to the slaughterhouse.

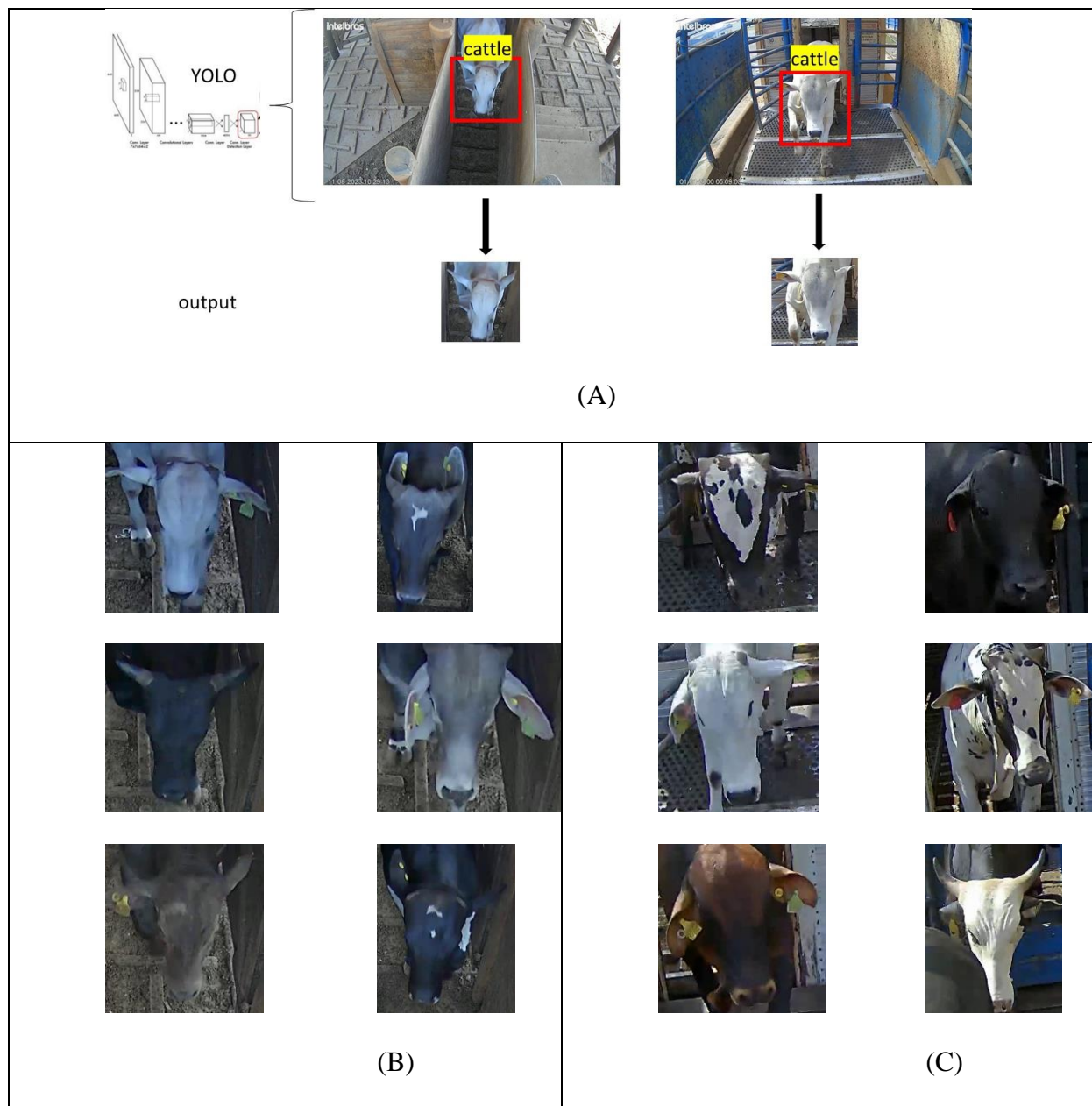


Figure 4.2 Example of Yolo network applied to videos of Location 1 and Location 2. (A) Face detection (B) Cropped images in Location 1. (C) Cropped images in Location 2.

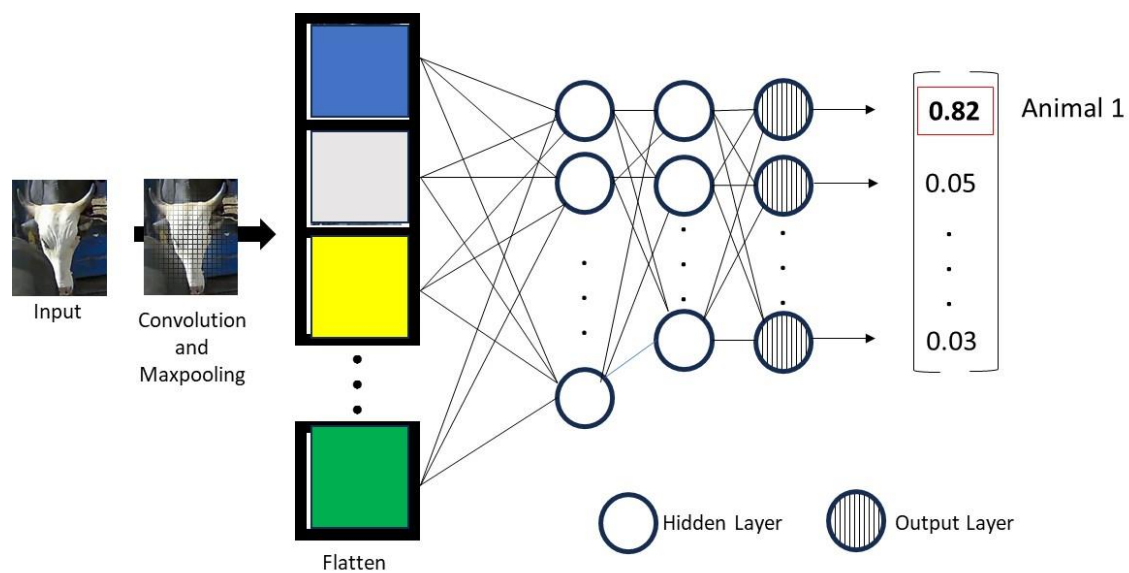


Figure 4.3 Overview of CNN architecture for beef cattle image identification in two locations of the livestock industry.



Figure 4.4 Examples of Nelore and Nelore crossbreed included in our models.

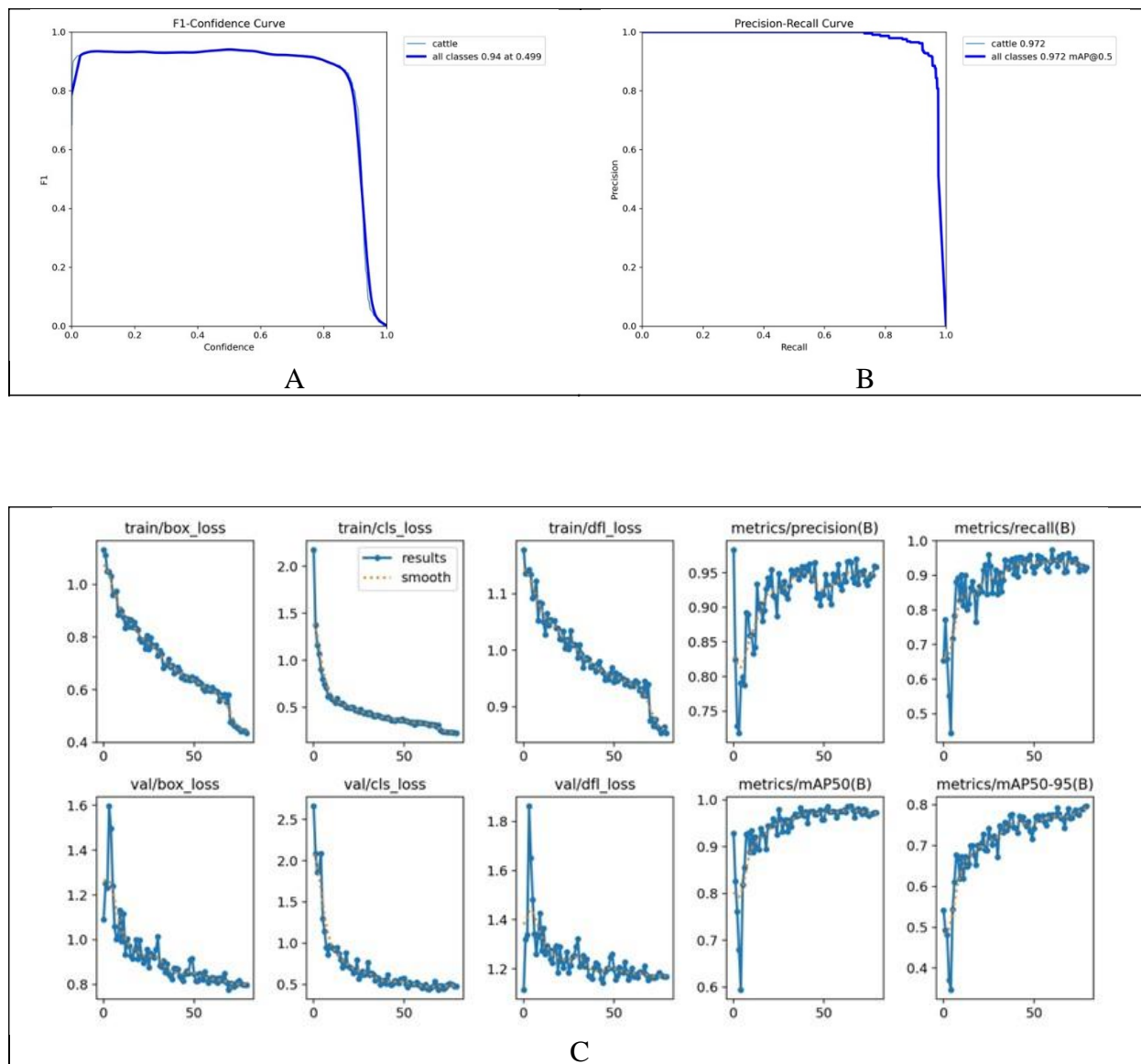


Figure 4.5 YOLOv8 performance in cattle face detection. (A) F1-confidence curve. (B) Precision-recall curve. (C) Loss and Mean Average Precision (mAP).

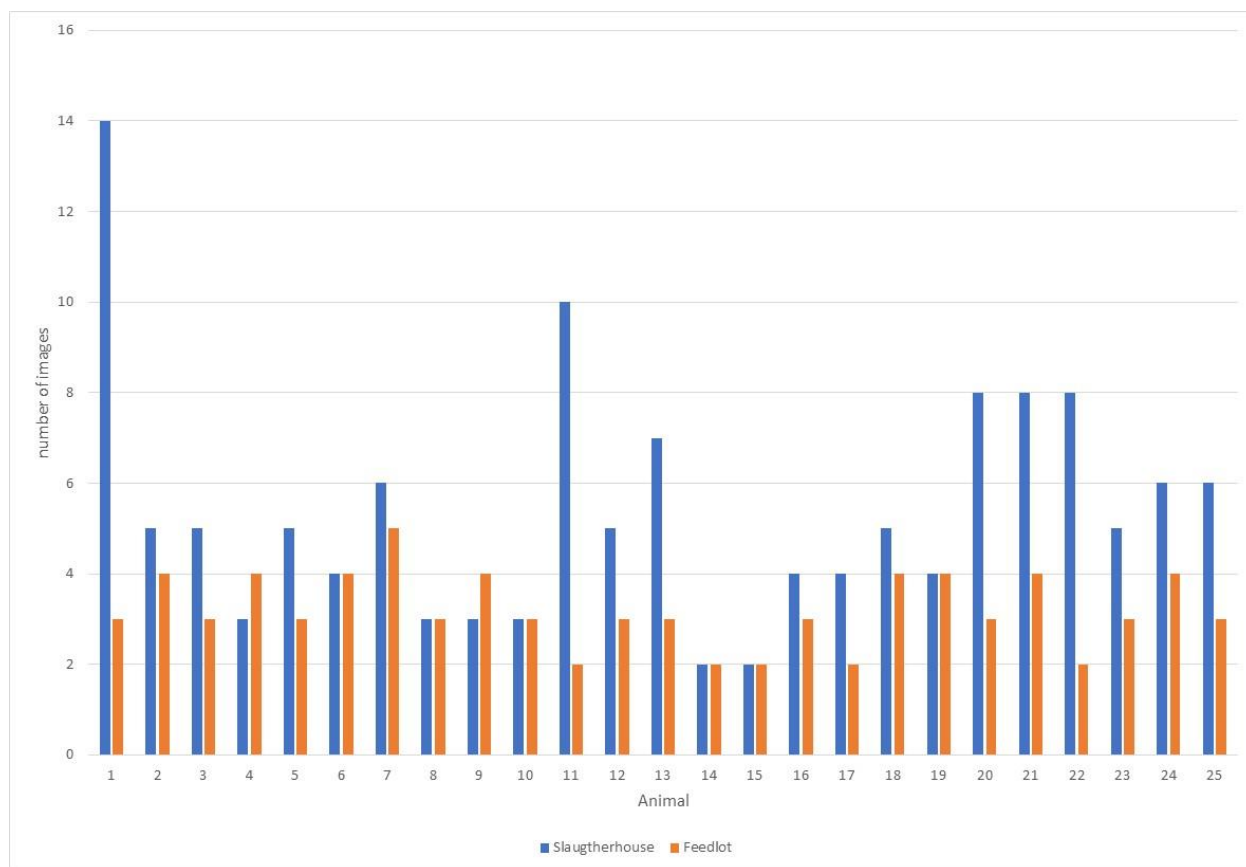


Figure 4.6 Frequency distribution of detected image's face per animal.

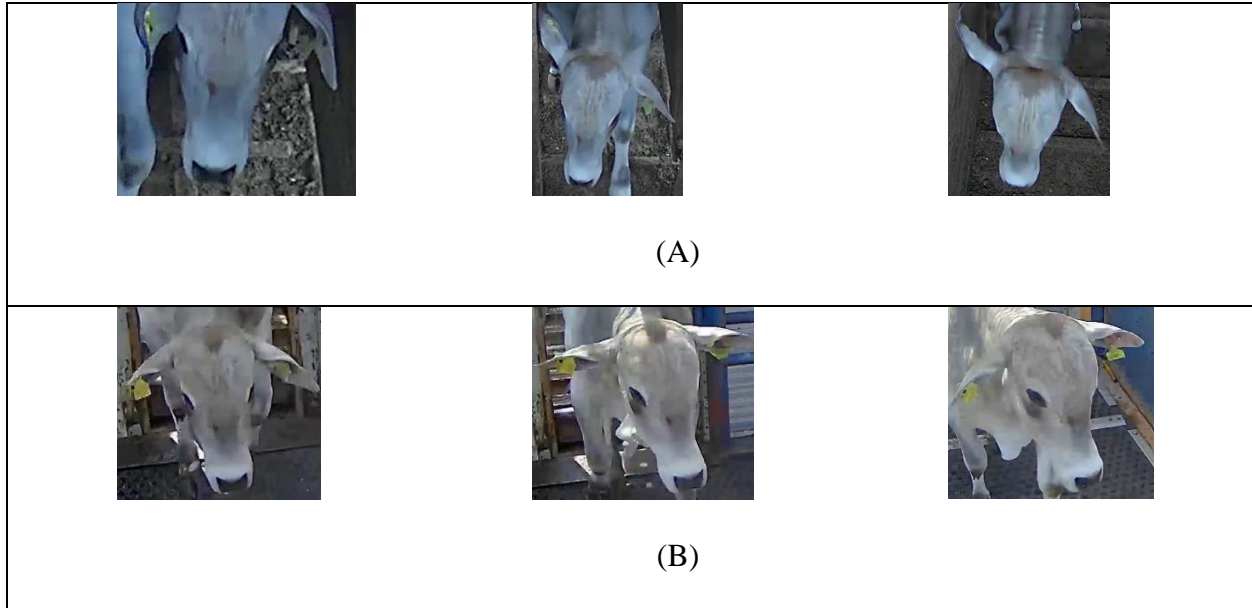


Figure 4.7 Example of repeated images of faces of the same animal in the experiment.
(A) Location 1. (B) Location 2.

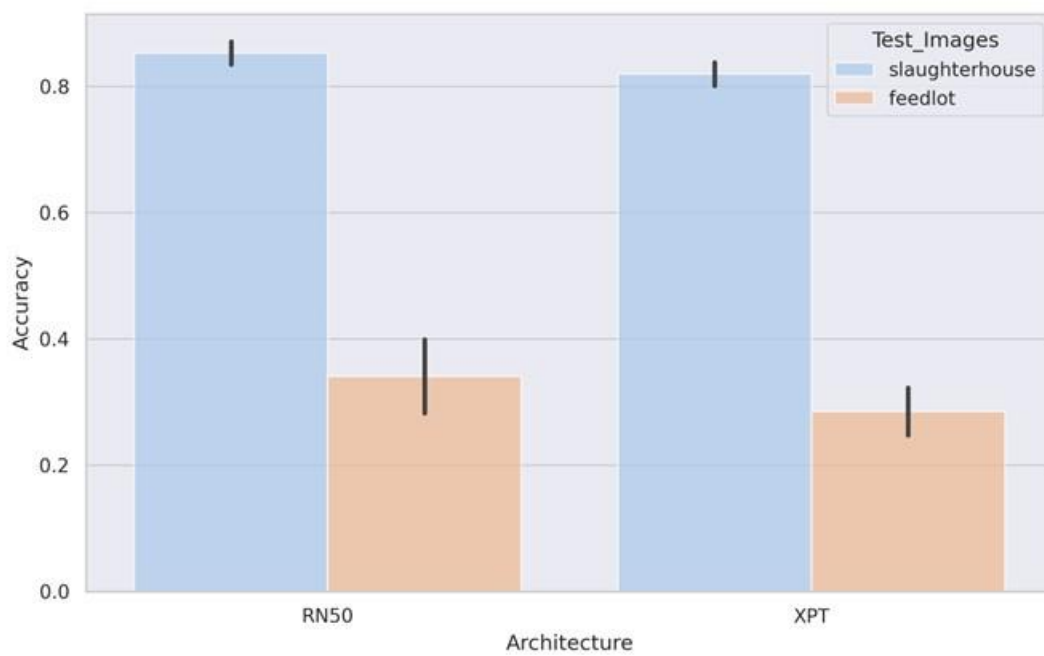


Figure 4.8 Accuracy of RN50 and XPT for animal classification considering the same location (slaughterhouse) for both training and test image sets and considering images generated at a different location (feedlot) as testing set.

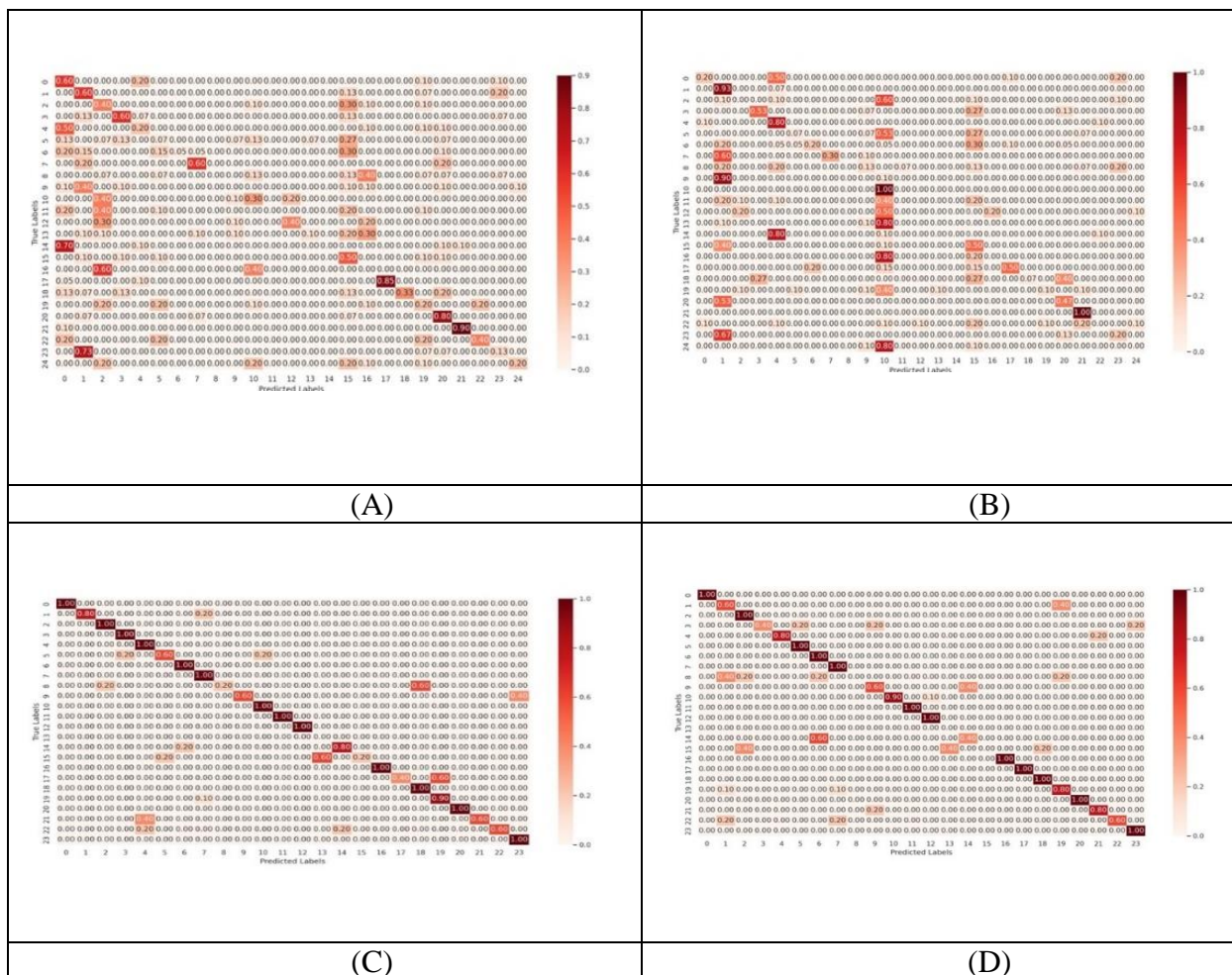


Figure 4.9 Heatmaps of confusion matrices for animal identification using location 2 (slaughterhouse) data as the training set. (A) RN50 with test set at location 1 (feedlot). (B) XPT with test set at location 1 (feedlot). (C) RN50 with Test Set at Location 2 (Slaughterhouse). (D) XPT with Test Set at Location 2 (Slaughterhouse).

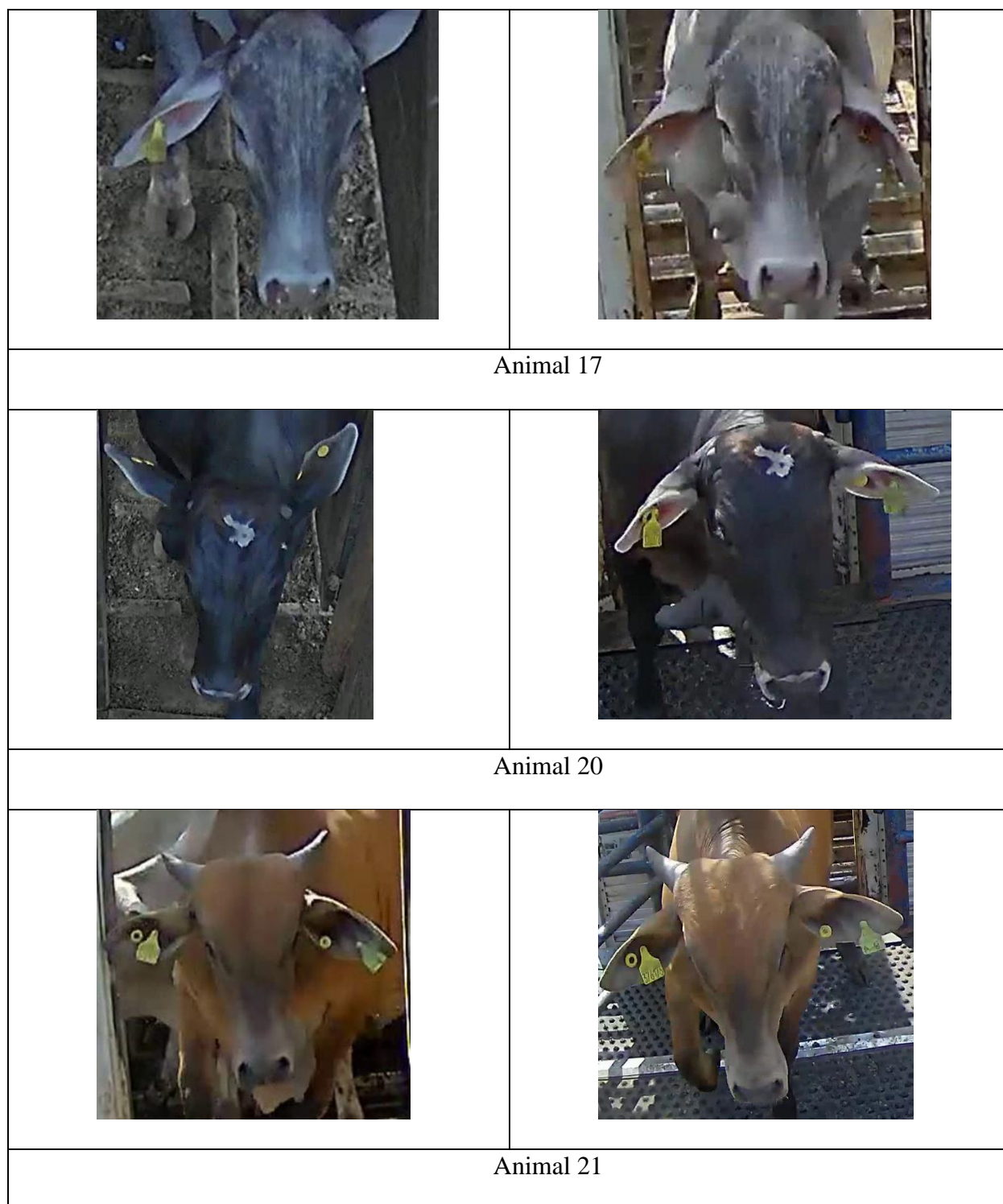


Figure 4.10 Examples of animals accurately identified at Location 1 when trained with images from Location 2 using the RN50 model.

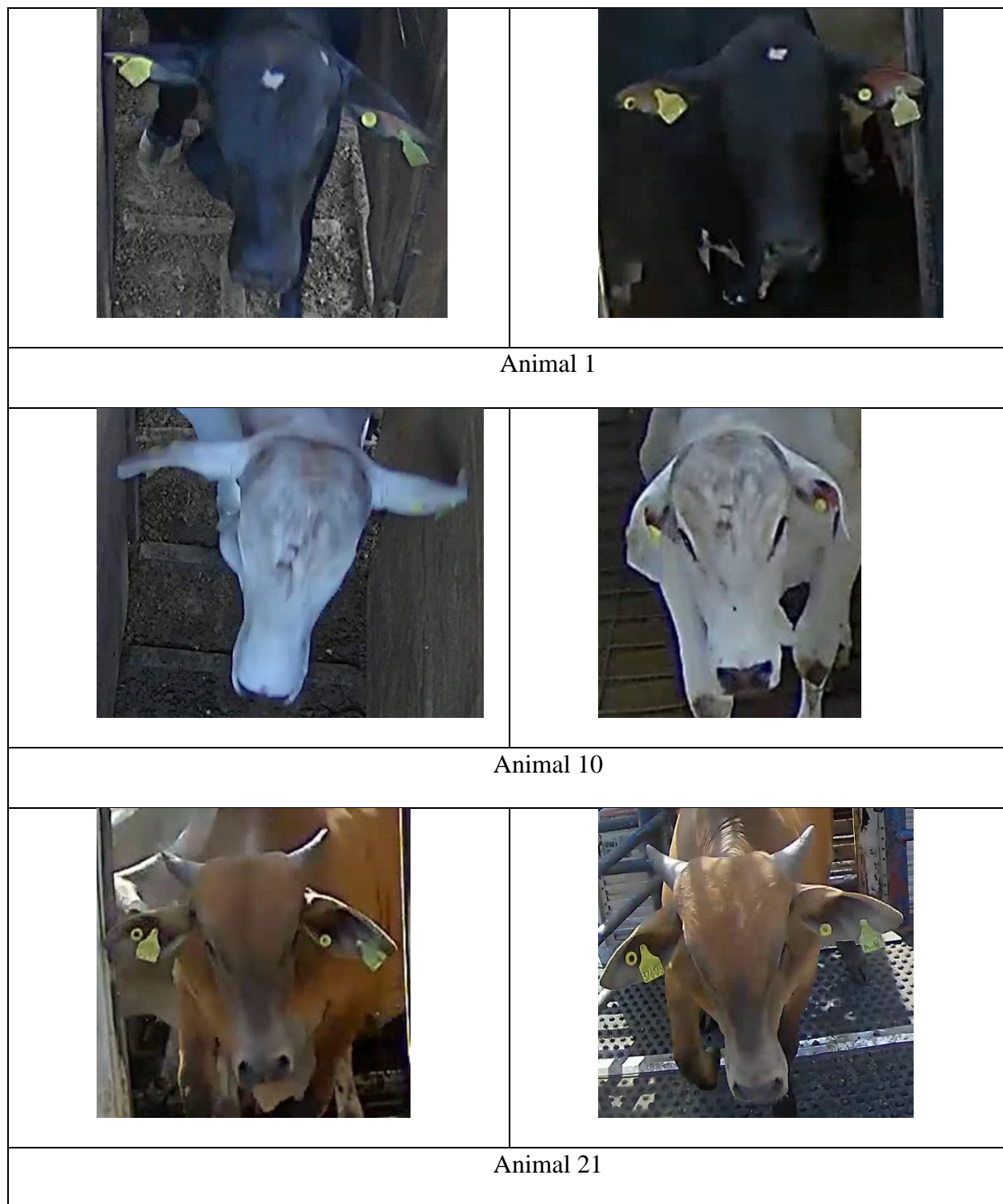


Figure 4.11 Examples of animals accurately identified at Location 1 when trained with images from Location 2 using the XPT model.

References

- Baber, J. R., J. E. Sawyer, and T. A. Wickersham. 2018. Development and comparison of methodology for estimating beef cattle contribution to human food supply 2. Protein quality and net protein contribution. *Anim. Sci.* 96:60.
- Bezen, R., Y. Edan, I. Halachmi. 2020. Computer vision system for measuring individual cow feed intake using RGB-D camera and deep learning algorithms. *Comput. Electron. Agric.* 172:105345. doi:10.1016/j.compag.2020.105345.
- Borges Oliveira D. A., L. G. Ribeiro Pereira, T. Bresolin, R. E. Pontes Ferreira, and J. R. R. Dorea. 2021. A review of deep learning algorithms for computer vision systems in livestock. *Livest. Sci.* 253:104700. doi:10.1016/j.livsci.2021.104700.
- Bromley, J., I. Guyon, Y. Le Cun, E. Säcker, and R. Shah. 1993. Signature verification using a “Siamese” time delay neural network. *Proc. 6th International Conference on Neural Information Processing Systems (NIPS’93)*. 737:744.
- Buslaev, A., V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin. 2020. Albumentations: Fast and flexible image augmentations. *Information* 11:125. doi:10.3390/info11020125
- Chollet, F. 2017. Xception: Deep learning with depthwise separable convolutions. *Proc. Computer Vision and Pattern Recognition (CVPR)*. 1251:1258.
- Cowton J., I. Kyriazakis and J. Bacardit. 2019. Automated individual pig localisation, tracking and behaviour metric extraction using deep learning. *IEEE Access*. 7:108049-108060. doi:10.1109/ACCESS.2019.2933060.
- Diwan, T., G. Anirudh, and J. V. Tembhurne. 2023. Object detection using YOLO: challenges, architectural successors, datasets and applications. *Multimed. Tools* 82:9243–9275. doi:10.1007/s11042-022-13644-y.
- Figueroa-Mata G., and E. Mata-Montero. 2020. Using a convolutional Siamese network for image-based plant species identification with small datasets. *Biomimetics*. 5:8. doi:10.3390/biomimetics5010008.
- Fu, Y., Y. Lei, T. Wang, W. J. Curran, T. Liu, and X. Yang. 2019. Deep learning in medical image registration: A review. *Med. Phys.* 46: 2037-2050. doi:10.1002/mp.13451.
- Hansen, M. F., M. L. Smith, L. N. Smith, M. G. Salter, E. M. Baxter, M. Farish, and B. Grieve. 2018. Towards on-farm pig face recognition using convolutional neural networks. *Comput. Ind.* 98:145–152. doi:10.1016/j.compind.2018.02.016.
- Haskins, G., U. Kruger, and P. Yan. 2020. Deep learning in medical image registration: A survey. *Med. Image Anal.* 58:101537. doi:10.1016/j.media.2019.101537.

- He, K., X. Zhang, S. Ren, and J. Sun. 2016. Deep residual learning for image recognition. *Proc. Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. 1:770-778.
- Hill, D. L., P. G. Batchelor, M. Holden, and D. J. Hawkes. 2001. Medical image registration. *Phys. Med. Biol.* 46. doi:10.1088/0031-9155/46/3/201.
- Huhtala, A. 2007. Evaluation of instrumentation for cow positioning and tracking indoors. *Biosyst. Eng.* 96:399–405. doi:10.1016/j.biosystemseng.2006.11.013.
- Islam, M. N., J. Yoder, A. Nasiri, R. T. Burns, and H. Gan. 2023. Analysis of the drinking behavior of beef cattle using computer vision. *Animals*. 13:2984. doi:10.3390/ani13182984.
- Kumar, S., A. Pandey, K. Sai Ram Satwik, S. Kumar, S. K. Singh, A. K. Singh, and A. Mohan. 2018. Deep learning framework for recognition of cattle using muzzle point image pattern. *Measurement* 116:1–17. doi:10.1016/j.measurement.2017.10.064.
- Lee T., Y. Na, B. G. Kim, S. Lee, and Y. Choi. 2023. Identification of individual Hanwoo cattle by muzzle pattern images through deep learning. *Animals* 13:2856. doi:10.3390/ani13182856.
- Li, G., G. E. Erickson, and Y. Xiong. 2022. Individual beef cattle identification using muzzle images and deep learning techniques. *Animals* 12:1453. doi:10.3390/ani12111453.
- Marsot, M., J. Mei, X. Shan, L. Ye, P. Feng, X. Yan, C. Li, and Y. Zhao. 2020. An adaptive pig face recognition approach using convolutional neural networks. *Comput. Electron. Agric.* 173: 105386. <https://doi.org/10.1016/j.compag.2020.105386>.
- McFarlane, N. J. B., and C. P. Schofield. 1995. Segmentation and tracking of piglets in images. *Machine Vis. Apps.* 8:187–193. doi:10.1007/BF01215814.
- Ni L. M., Y. Liu, Y. C. Lau and A. P. Patil. 2004. LANDMARC: indoor location sensing using active RFID. *Proc. International Conference on Pervasive Computing and Communications* 1: 407-415. doi: 10.1109/PERCOM.2003.1192765.
- OECD/FAO. 2023. OECD-FAO Agricultural Outlook 2023-203. OECD Publishing. doi:10.1787/08801ab7-em.
- Qiao, Y., D. Su, H. Kong, S. Sukkarieh, S. Lomax, and C. Clark. 2019. Individual cattle identification using a deep learning based framework. *IFAC-PapersOnLine*. 52:318–323. doi:10.1016/j.ifacol.2019.12.558.
- Shao D., Z. He, H. Fan, and K. Sun. 2023. Detection of Cattle Key Parts Based on the Improved Yolov5 Algorithm. *Agriculture*13:1110. doi:10.3390/agriculture13061110.

- Sergeant D., R. Boyle, and M. Forbes. 1998. Computer visual tracking of poultry. *Comput. Electron. Agric.* 21:1-18. doi:10.1016/S0168-1699(98)00025-8.
- Tan K. G., A. R. Wasif and C. P. Tan. 2008. Objects tracking utilizing square grid RFID reader antenna network. *J. Electromagn. Waves Appl.* 22:27-38. doi:10.1163/156939308783122724.
- Tillett R. D., C. M. Onyango, and J. A. Marchant. 1997. Using model-based image processing to track animal movements. *Comput. Electron. Agric.* 17:249-261. doi:10.1016/S0168-1699(96)01308-7.
- Zhao, K., He, D. 2015. Recognition of individual dairy cattle based on convolutional neural networks. *Trans. Chin. Soc. Agric. Eng.* 1:1002-6819. doi:10.3969/j.issn.1002-6819.2015.05.026.
- Zheng, S., S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. S. Torr. 2015. Conditional random fields as recurrent neural networks. *Proc. International Conference on Computer Vision (ICCV)*.1:1529–1537. doi:10.48550/arXiv.1502.03240.
- Zitova, B., and J. Flusser. 2003. Image registration methods: a survey. *Image Vis. Comput.* 21: 977-1000. doi:10.1016/S0262-8856(03)00137-9.

CHAPTER FIVE: CONCLUDING REMARKS

This dissertation explored some applications of computer vision to beef cattle production. More specifically, this dissertation focuses on: 1) Developing a comprehensive image analysis pipeline to process top-view images of beef cattle in outdoor feedlot environments for subsequent animal weight predictions, 2) Creating an additional module within the mentioned pipeline to predict cattle live weight using features extracted from the animals' body images, and 3) Investigating methods for identification of animals within and across two distinct locations in the cattle industry using facial images.

The application of computer vision systems (CVS) to outdoor feedlot environments, characterized by the presence of freely moving animals, multiple animals within the same scene, and various light interferences, presents a significant technical challenge. To overcome these obstacles, Chapter 1 introduces a processing pipeline comprising four fundamental modules for the analysis of top-view images of beef cattle in feedlots aiming to extract features of the animal's body, such as dimensions and shape, for body weight (BW) predictions. The first module enhanced image quality by combining infrared and depth imaging techniques, reducing image noise, and highlighting the animals in the scenes. The second module identified the most effective approach for automatically classifying usable images, i.e., those images containing one single full-body animal within the scene, while discarding unusable ones (i.e., those images with multiple animals, only partial animal body, or empty scenes). The AlexNet neural network, when used with the infrared and depth composed image, consistently achieved the highest level of accuracy. The third module improved the classifier performance using data augmentation with synthetic images. Finally, the fourth module successfully employed Marker-Controlled Watershed Segmentation (MCWS) to separate the animal's body from its head and tail on each image. These steps

collectively lead to the acquisition and processing of images exclusively containing the animal's body, thus preparing them for subsequent BW predictions.

Upon obtaining animal body images, the analysis proceeded to extract body measurements for BW prediction. However, challenges in outdoor feedlot environments, such as variable animal positions, fluctuating light, temperature, and other complexities, compromised image-derived feature accuracy, affecting precision of BW prediction. To address these challenges, in Chapter 2 we compared Fourier feature extraction versus traditional image-derived features and also evaluated four machine learning (ML) techniques (partial least squares regression, ridge regression, lasso regression, and artificial neural networks) for BW prediction. Consistently, Fourier transformation, especially when combined with artificial neural networks, outperformed other modeling approaches.

A thorough understanding of animal origin, movement, and individual beef cattle identification within the meat production chain is vital for ensuring product quality and safety, preventing disease outbreaks, optimizing logistics efficiency in cattle transportation, and complying with industry regulation. In Chapter 3 we investigated CVS approaches for animal identification by conducting a preliminary study to explore the potential of traditional neural networks in identifying animals within and across two distinct locations along the meat production chain using their facial images. Our results showed that conventional CNN-based models achieved reasonable accuracy when tested within the same location but faced challenges with cross-location testing, resulting in reduced accuracy. The comparison of the confusion matrices derived from two distinct network architectures, *RN50* and *XPT*, allowed us to discern their individual strengths. *RN50* excelled in detail extraction, such as detecting coat color spots and horns. Conversely, *XPT* demonstrated proficiency in capturing general features and effectively managing variations in

lighting and positioning, particularly in scenarios with heterogeneous conditions. This study demonstrated that we can prospect networks capable of encompassing both abilities or contemplate them by combining these two neural networks to improve our results. Additionally, this study highlighted the importance of aligning angles in images captured at various locations throughout the animals' journey and underscored the need of utilizing neural networks capable of handling relatively small samples, such as Siamese networks.

Overall, this dissertation has achieved its goal by presenting advanced techniques in image processing, along with state-of-the-art ML models tailored for real-world farm and industry conditions. These tools can be potentially combined to develop a system capable of automatically identifying animals and monitoring their weight without need to restrict the animals to a specific space, minimizing stress on them, and eliminating the requirement for human intervention in image acquisition. Furthermore, it is essential to underscore the importance of exploring image processing techniques rather than applying ML methods such as neural networks indiscriminately. It is not always suitable to depend solely on artificial intelligence algorithms, as other image processing techniques may complement or even prove more effective in specific contexts. The field of image processing offers a wealth of theories and concepts that can be highly beneficial in the realm of animal husbandry, especially in challenging real-world conditions, as addressed in this study.