Computer vision applied to automated measurement of biometric traits and prediction of body weight and carcass traits in live animals

by

Arthur Francisco Araujo Fernandes

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Animal Sciences)

at the

University of Wisconsin-Madison

2019

Date of final oral examination: 07/23/2019

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

Dr. Guilherme J. M. Rosa, Professor, Animal Sciences & Biostatistics

Dr. Daniel Gianola, Professor, Animal Sciences & Dairy Science

Dr. Vikas Singh, Professor, Computer Sciences & Biostatistics

Dr. David L. Thomas, Professor, Animal Sciences

Dr. Kent Weigel, Professor, Dairy Science

Acknowledgments

I would like to express my eternal gratitude to my advisor Dr. Guilherme J. M. Rosa, who accepted me as his advisee with open arms and put his trust in my crazy ideas (maybe not too crazy…). Without your guidance and experience, I do not know which turn my doctorate would have taken, and I am extremely glad about what we have accomplished, inside and outside of this dissertation.

I also wanted to acknowledge the members of my committee, Dr. Daniel Gianola, Dr. David Thomas, Dr. Vikas Singh, and Dr. Kent Weigel for their continuous advice and support through the time here Madison. I was very happy when each of you accepted to be a member of my committee, at a time where I was unsure of which directions this dissertation would take. My special thanks to Dr. Vikas Singh for the advice and discussions about computer vision and image analysis in the moments of need and for your great suggestions of materials to study. To Dr. David Thomas who taught me how to teach, it was my pleasure to assist in your lectures and to be able to learn from you. To me, you are a great example of a good Professor and an inspiration. I thank Dr. Daniel Gianola for being an example of a researcher, always with a curious mind and a pertinent question about virtually anything, one of my regrets was to not be able to take your course. Finally, I would like to express my thanks to Dr. Kent Weigel for his continuous support to my research, and for the valuable feedback.

I also want to thank the Brazilian government (CAPES agency) for the financial support of my study. I would like to thank Dr. Eduardo Turra, my friend and previous advisor, for gladly providing part of the data used in this dissertation. I would like to thank the people from Genus PIC, especially to Dr. Robert Fitzgerald for believing in this

research and supporting it. This dissertation would have never existed without this support.

I would like to express my gratitude to the faculty and staff members of the Animal Sciences Department, the Writing Center and UW Health for all their support. In special to Minh, Steve, Kathy, and Deb for helping me with the most basic, but also crucial tasks of the graduate student life, and for the diligence in doing so. I cannot express enough my gratitude; I consider you as my friends.

This acknowledgment would never be complete if I do not express my thanks to my family and friends. Without their support, I would have never built the strength and courage to pursue a doctorate, nor I would have had the tenacity to complete it. Mostly my wife Anna, for pushing me to always be my best self, and who is also crazy enough to pursue a doctorate at the same time. You are my best friend and the source of my strength. To my parents, you are always the great heroes of my life, my sisters and my godmother who came to visit me and gave me a so much needed boost of energy. Thank you for believing in me and for all the mixture of talks, discussions, fights, affection, and laughs that helped me to stay true. To the whole Araújo and Fernandes family, I wish I could list each and every one, I tried, love you all. To my friends that are also my family, Luis, Rafaela, Lourenço, Norma and David. To the friends that I left in Brazil, especially Lucas, Marcos, Samuel, Eduardo Slot, Leo, for the talks that helped reducing the distance. To my friends Tiago and Rodrigo who were with me from the beginning of this journey. To the friends from Madison, especially to Renata, Ligia Moreira, Vera, Bruno, Vivian, Fernando, Paulo, Angelina, Isabela, Gerson, Lucio, Samuel, Nora, João, Tom, Tiago, and Camila. You all helped to make these last years a wonderful roller-coaster.

Contents

List of Tables

List of Figures

Abstract

This dissertation deals with applications of computer vision in animal sciences. The first chapter provides an introduction to computer vision and literature review of applications in animal sciences. The second chapter focuses on digital image analysis for Nile tilapia in which images are used for extraction of body measurements and prediction of body and carcass weight, as well as carcass yield. The best segmentation network achieved results for intersection over union of 99, 90 and 64% for background, fish body and fin areas, respectively. Also, a predictive model including body area as covariate achieved R^2 of 0.95 and 0.94 for body and carcass weights, respectively. However, none of the evaluated models gave satisfactory predictions of carcass yield. The third chapter aims at the use of 3D images for measurement of pig biometric traits and body weight. In this chapter, classical image processing strategies were used for the development of a system for classification of a pig image, extraction of variables, and prediction of body weight. A linear model using the truncated median estimate of selected variables achieved an R^2 of 0.94 for prediction of body weight. This result was similar to previous reports in which a human evaluator manually selected images from animals to be used. In conclusion, it was possible to achieve satisfactory results for prediction of pig body weight automatically using 3D sensors. In the fourth chapter, several modeling approaches were evaluated for prediction of body weight, lean meat, and back fat in pigs. Again, 3D images were processed for extraction of image variables that were used as input for either linear regression and machine learning models. In addition, a deep learning approach using the raw image as input was evaluated. Over the evaluated

models, the deep learning approach presented the best predictive performance for body weight and muscle depth, while for back fat, there was no clear advantage compared to linear regression. In conclusion, it was possible to predict body weight, muscle depth, and back fat using 3D images of the pigs. Nonetheless, additional research should be performed, towards the improvement of predictions of back fat.

Chapter 1: Review on Computer Vision and applications to Animal Sciences

1.1 Introduction

As sighted human beings, we experience vision in a way that seems natural and automatic. From the moment we are born we use our vision system to navigate the world around us and to identify and interact with other beings and objects. Therefore, this vision system is constantly training and adapting so that we are capable of using it for several tasks. The system works with the luminous signal being captured by the eye, transferred via the optic nerve to the brain where it is processed and interpreted (Gregory, 1978). This complex vision system can adapt to different light conditions autonomously while allowing us to focus on objects and to have a 3-dimensional representation of the world. But what would be vision for a computer? This chapter is divided into two sections, the first section provides a brief introduction to computer vision, describing current developments and example applications. The second section focuses on computer vision applications in animal sciences describing different technologies, as well as, common sensors and algorithms used. This chapter closes by presenting areas of current interest and future directions for computer vision applications in animal sciences.

1.2 What is Computer Vision?

Computer vision is the field that tries to describe the world through images, interpreting or reconstructing properties of the world that can be extracted from these

images, such as shapes, textures, density, and distance (Szeliski, 2010). Computer vision systems (CVS) are also known as machine vision systems, visual image systems, or just image systems. Therefore, computer vision essentially is the development of artificial systems to handle visual problems of interest. However, there are other terms used to describe the same field and interconnected areas, such as digital image processing, digital image analysis and pattern recognition, among others.

In short, *Digital image processing* can be viewed as the study of obtaining a visual signal of the world and processing it in order to make it interpretable. It spams from the study of image formation, as a result of the acquisition of light signals by specifically designed sensors, to the interpretation of the image as an array of connected values. Therefore, digital image processing involves the conception, design, development, and enhancement of digital imaging algorithms and programs (Burger and Burge, 2016). As such, it is a discipline heavily based on physics and mathematics. The term can also be used to directly address the applications or techniques used for digital image manipulation, ranging from noise reduction, image equalization, image filtering and other transformations used for preparing images for subsequent steps in an analysis pipeline or for enhancing images aesthetically.

Image analysis, on the other hand, corresponds to the process of extracting meaningful information from the image. This information can be global image metrics, as colors histogram and brightness distribution, or block statistics, as intensity, moments (mean, variance) and integral images. This information extracted from the image can be used for imaging processing techniques, such as sharpening, thresholding and smoothing (Otsu, 1979; Polesel et al., 2000; Rosin and Ioannidis, 2003; Kheradmand and Milanfar,

2015). However, image processing techniques can be applied also prior to image analysis techniques, such as image segmentation. There are several methods of image segmentation, but basically they can be classified into methods that do a global clustering of image pixels according to some criteria independent of spatial information; methods that account for local information, such as texture, color, edges and integral images (Rother et al., 2004; Nakagomi et al., 2013; Tang et al., 2013).

Pattern recognition is another field that is also intimately related to image analysis, and it is a field that studies not only images but other signals, such as sound and texts. As the name suggests, it is an area dedicated to the study of patterns that may appear in any given signal. In the context of imaging, pattern recognition is the development of mathematical methods for identification of simple geometrical structures as lines or key point features that can be jointly used to identify more complex objects or patterns (Hough, 1962; Rosten and Drummond, 2005; Bay et al., 2006; Calonder et al., 2010; Leutenegger et al., 2011).

Before an image could be processed and analyzed in a computer, there was the need to develop sensors able to recognize, measure and digitalize luminous signals. It was in the 1970's, with the advent of charge coupled devices (CCD) sensors (Boyle and Smith, 1973), that digital imaging was developed, and the interest in computer vision systems appeared. The digital signal captured by the sensor is coded and stored in arrays of data that can be interpreted and manipulated by computational algorithms. Thus, for a computer, an image is nothing more than numerical values in a structured array of data that codifies light and colors for each point in the image. This array can be a single matrix, for grayscale images, where the values inside the matrix correspond to different shades

of gray or an array of 3 matrices in the case of color images (i.e. intensities of red, green and blue, on the RGB color space) or even multiple matrices for hyperspectral images. Therefore, mathematical manipulations and statistics of an image were among the first studies developed in digital image analysis and processing.

Another turning point in the history of computer vision was the advent of personal digital cameras in the 1990's, reducing the costs and popularizing the process of capturing and analyzing digital images (Lister, 1995; Pratt, 2007; van Dijck, 2008). Since then, several applications of digital photography appeared. This popularization of digital cameras is directly connected to the increasing volume of data (photos and videos) generated over the last few years in many areas which are attempting to use computer vision to solve the most diverse problems. This increased interest in computer vision and related areas can be illustrated by the increasing content of scientific knowledge produced in the last decade (Figure 1.1).

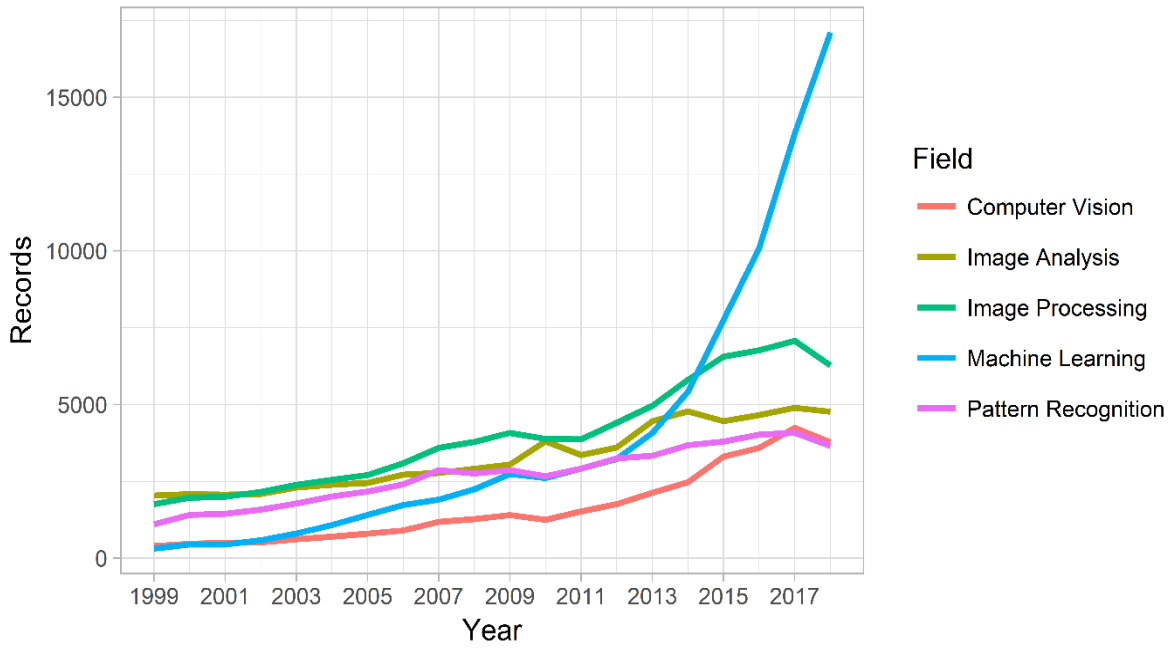

Figure 1.1 *Count of publications hits in "Web of Sciences" by related fields.*

Some areas of recent interest in computer vision are object sensing, mapping, recognition, motion tracking, navigation, image segmentation, and scene interpretation. However, while humans and animals do most of these actions intuitively, most of the vision tasks are considered as difficult problems in computer science, and the algorithms available are prone to errors. Thus, most of the successful CVSs are results of multidisciplinary approaches tailored for specific cases, such as interactive segmentation (Rother et al., 2004), face detection based on image features (Viola and Jones, 2004), and machine learning methods for object recognition and classification (Lecun et al., 1998; Litjens et al., 2016). For such tasks, CVS generally use algorithms and principles of pattern recognition, image analysis, and processing in order to tackle the most diverse problems. Figure 1.2 presents a possible framework of a CVS, which can be passive, with a fixed sensor capturing the information that is presented by the world or actively explore the world and adjust its perception (field of view, exposure, among others).

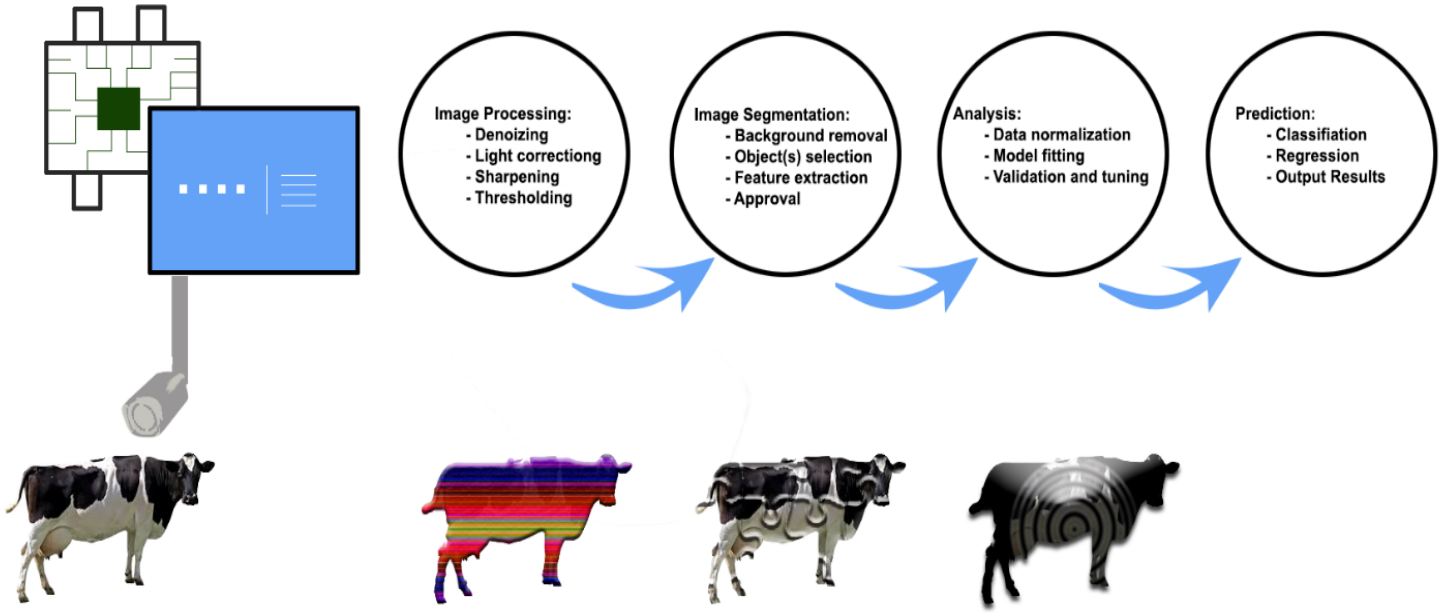

Figure 1.2 *Example of a Computer Vision System framework.*

From the machine learning algorithms used in computer vision, it is worth mentioning Deep Learning algorithms, which have recently been successfully used in diverse vision applications. These algorithms are a type of machine learning that is an extension of traditional artificial neural networks (ANN) that achieves great power and flexibility by learning more abstract representations of the inputs as a nested hierarchy of concepts (Goodfellow et al., 2016). These nested concepts, or hidden layers, generate very complex models with many parameters that were possible to be trained only with the advent of very large datasets, data augmentation techniques and advancements in ANN, such as, the development of learning optimization via stochastic gradient descending, new activation functions such as ReLU, regularization techniques, and efficient use of GPUs (Krizhevsky et al., 2012; Goodfellow et al., 2013; LeCun et al., 2015).

1.3 Applications of Computer Vision Systems in Animal Sciences

Before the advent of CVS, many applications in animal sciences required the use of the trained eye for visual classification of live animal behavior, body condition score, carcass fat deposition, meat marbling, and classification of eggshell quality. There are also methods that use the aid of lenses, such as microscopes, for evaluation of cell morphology in a blood smear or spermatozoid motility and defects. Moreover, other signals such as ultrasound, infrared and x-rays are widely used to produce images for diagnostic purpose. However, most of the methods currently used for the measurement of traits of interests need expert personnel. This requires training of evaluators from time to time in order to maintain good measurement quality. Also, most of the measuring process is time demanding and stressful to the animals and costly for the farmer, making it prohibitive due to animal welfare and economic reasons. Therefore, there is an interest

in developing automatic, indirect methods for monitoring livestock and measuring traits of interest. In this section, we describe the applications of CVSs in animal sciences as an answer to the need for such automatic, non-invasive methods. The section is divided into methods for carcass and meat traits and for live animal monitoring and phenotyping as there are different challenges involved for each.

1.3.1 Carcass and Meat traits

Probably one of the earliest applications of a CVS was in meat sciences with the earliest reported studies found in the 1980's (Lin, 1978; Cross et al., 1983; Wassenberg et al., 1986). In these studies, the system was composed of a camera, light source, digitizer, and computer unit. The CVS needed an operator to position beef meat cuts on a surface at a known distance and angle from the camera and to trigger the image acquisition. Thus, the meat cuts were all positioned in the same manner with constant background and illumination. At that time, the interest was to predict the cut content of lean meat and fat and to compare the results from the CVS to trained USDA meat graders. Even though the system was not fully automated, the prediction equations developed with the traits measured by the system presented better results than the prediction equations developed with traits measured by trained graders. The best model for prediction of lean meat weight developed with the CVS traits achieved an R^2 of 0.93 against 0.84 from the model that included carcass traits and the USDA marbling score as covariates.

Since these earlier studies, there has been an increasing interest in the use of computer vision for prediction of the most diverse meat quality traits, not only for beef but also for fish, poultry, and pork (Table 1.1). Many recent studies are focused not only on the crude content of meat and fat but also on more refined chemical characteristics (Tao

and Ngadi, 2017) and prediction of meat quality classification, palatability, tenderness, and other traits normally evaluated by a panel of trained experts (Jackman et al., 2010; Jackman et al., 2011; Zapotoczny et al., 2016). Currently, the most commonly used devices in this area are color digital cameras as a CCD sensor alone, or with a complementary metal-oxide-semiconductor (CMOS) sensor. Other technologies that have also been used in meat sciences for image acquisition are ultrasound, infra-red, ultraviolet, hyperspectral imaging, and, magnetic resonance (Jackman et al., 2011). These alternatives are generally more expensive than standard digital cameras, however, each different imaging technology can be used for specific applications, such as evaluation of meat internal attributes and chemical characteristics (Xiong et al., 2014; Tao and Ngadi, 2017).

1.3.2 Monitoring and Phenotyping of Live Animals

One of the most desired applications of CVS for live animals is in monitoring and identifying changes in their behavior. Animals tend to have a common behavior within a group (e.g. animals of the same species/breed and physiological stage), and changes on their behavior may be caused by environmental changes, management problems or disease. There are also differences between animals inside the same group, as there are dominant and passive behaviors. Therefore, there is a constant effort to understand behavioral changes and their relationship with other traits of interests. Routine evaluation of the animal is normally conducted by trained evaluators at specific time points, such as at the time of transfer from one location to another (e.g., from nursery to grow-out farms) and vaccination. This is because managers and workers have limited time to observe a

group of animals, and there is an increase in the animal/manager ratio. Thus, the most common use of CVS for evaluation of animal behavior is for acquisition and storage of images and videos that can be assessed later or remotely. This improves animal management since there is no need for the evaluator to be present, which otherwise can cause behavioral changes on the animals. Also, the evaluator can loop across images, and replay then to improve the quality of the evaluation. Nonetheless, it is still desirable to have a system to store the images and also automatically classify the behavior and alert the manager of important changes.

Table 1.1 *Examples of image computer vision applications in meat sciences*

	Applications	**Image signal**	**References**
Cattle and Small Ruminants	Carcass	D; US; VL	(Stanford et al., 1998; Cannell et al., 1999; Teixeira et al., 2008; Gomes et al., 2016)
	Classification	VL	(Shackelford et al., 2003; Steiner et al., 2003)
	Fat (kg and%)	US; VL	(Cross et al., 1983; Yang et al., 2006; Teixeira et al., 2008; Nunes et al., 2015)
	Lean meat (kg and %)	VL	(Cross et al., 1983; Cannell et al., 1999)
	Marbling	VL	(Shackelford et al., 2003; Steiner et al., 2003)
	Tenderness	VL	(Vote et al., 2003; Jackman et al., 2010; Nunes et al., 2015)
Fishery	Fat Pigmentation	IR; VL	(Folkestad et al., 2008)
	Quality	IR; VL	(Kimiya et al., 2013)
Poultry	Classification	H; VL	(Zapotoczny et al., 2016)
Pork	Carcass	US; VL	(Smith et al., 1992; Doeschl-Wilson et al., 2005)
	Classification	H;VL	(Zapotoczny et al., 2016)
	Quality	H; IR; VL	(Wang et al., 2013)

D - Depth; H - Hyperspectral; I - Infrared; MR - Magnetic Resonance; T- Thermal; US - Ultrasound; UV - Ultraviolet; VL - Visible Light;

Differently from carcass and meat cuts that can be easily positioned for image acquisition under a well-controlled light source and even background, several obstacles arise when working with live animals. As an example, in farm conditions, the illumination can change throughout the day even inside a barn due to sun position, clouds and seasons. Moreover, there will be differences between artificial light sources from one farm/barn to another, as they may use different types of lamps with different voltage and positioning. The background is also going to be different in each location and is prone to changes across the seasons or across time for a given location. Examples of differences in the background are floor surface material for animals in a barn and vegetation for free range animals. Therefore, the diversity of situations is probably one of the biggest challenges in implementing CVS that are robust enough to present satisfactory performance across different farm conditions.

Initial works with pigs demonstrated the applicability of a CVS to identify the animal position and to track its movement (Tillett et al., 1997; Lind et al., 2005). Those works showed that the algorithms available at the time were able to segment a single pig from the background under specific conditions. The conditions were: camera positioned to get the top view of the animal and dark background for a white pig. The method developed by Tillett et al (1997) estimated a point distribution of landmarks on the pig contour for a sequence of frames and was able to model small changes in the animal's posture. However, it was prone to lose the animal if the changes in position were abrupt from one frame to another. On the other hand, Lind et al (2005) used a more robust segmentation approach based on the generation of a background matrix for image subtraction and consequently animal segmentation. Even though this method cannot identify animal

posture it was efficient for use in a real-time application and efficiently tracked differences in animal activity.

Traits related to animal activity are still of great focus in image analysis for animal science, and one of the biggest issues is still identification and tracking of multiple animals and their interactions. In order to overcome this challenge, a successful approach in pigs was to identify the animals by patterns printed on their back (Ahrendt et al., 2011; Kashiha et al., 2013b). By using this simple approach, researchers were able to track and identify multiple animals enabling the characterization and evaluation of simple activity status as active or non-active with high correlation (mean of 0.9) to evaluations made by a human observer (Kashiha et al., 2013c; Kashiha et al., 2014b; Ott et al., 2014). However, this identification and tracking approach cannot be used for animals with darker skin or in commercial farms that have animals with different skin colors since the method was developed for white pigs on a dark floor background and using a surveillance camera. Another challenge in the use of patterns printed on the animal's skin for identification in commercial settings is the high stocking density and pen size, where there are many more animals than in an experimental setting.

In order to identify more activity behaviors, like feeding and drinking behaviors, segmentation of the captured image in regions of interest (ROI) have proved effective (Mendes et al., 2011; Kashiha et al., 2013a; Nasirahmadi et al., 2017). The basic concept is to identify not only the animals but also objects, such as the water source and feeders, and track how animals interact with those objects. Machine learning techniques have proven efficient for identification of animal posture such as, standing, lying or sitting (Mattachini et al., 2013; Barnard et al., 2016; Lao et al., 2016; Nasirahmadi et al., 2017)

or interactions among animals, such as mounting and aggressive behavior, such as biting (Oczak et al., 2014; Viazzi et al., 2014b). Viazzi et al. (2014b) achieve mean accuracy of 0.88 when using linear discriminant analysis for classifying aggressive behavior in pigs, while Barnard et al. (2016) achieved a mean accuracy of 0.91 when using structural support vector machine to classify dog postures from depth images. However, supervised learning models need a dataset of labeled images (ground truth) for the training step. In order to produce those training datasets, a manual classification of the images by a human observer is needed, thus the model will be at most as good as the human observer that evaluated the images initially. In order to improve those models, crowdsourcing the development of the dataset, with the manual classification of the images been done by several evaluators, and the use of multiple model classifiers have promoted an increase in accuracy for applications in histology image analysis (Arganda-Carreras et al., 2015; Litjens et al., 2016).

Apart from cameras that acquire images from the visible light spectra, two other groups of sensors have been used in applications with live animals. The first group is that of depth sensors, whose images correspond to a distance mapping of the scene of interest in relation to the sensor. These images can be generated from a series of techniques such as binocular cameras and stereo mapping, structured light, and time of flight. The usage of this kind of data alone or jointly with digital color cameras can improve image segmentation and also measurements of area and volume of animals. Examples of successful applications include identification of landmarks on animal shape with consecutive modeling of gait (Jabbar et al.; Thorup et al., 2007; Stavrakakis et al., 2015; Salau et al., 2017), posture (Viazzi et al., 2014a) and body condition score (Spoliansky et

al., 2016). Studies in gait analysis usually demanded intense manual label of video frames by a human observer and/or expensive system of plate markers to be positioned on the animal body and multiple cameras (Stavrakakis et al., 2014). However, with the introduction of time of flight technology, CVS with a single or two sensors were capable to efficiently estimate walking kinematics in pigs in a cost-effective framework with prediction accuracy comparable to state of the art systems (R^2 = 0.99) (Stavrakakis et al., 2015). Another application of CVS based on depth sensors is for estimation of animal body measurements (heights, widths, area, and volume as examples) and body weight. Although it is possible to estimate animal weight with only a standard surveillance camera as shown by Kashiha et al. (2014a), and achieving good prediction (R^2 = 0.95), a CVS based on depth image can extract additional information on animal volume and, thus achieve an R^2 of 0.99 (Kongsro, 2014) on similar experimental conditions.

The second group of sensors is composed by cameras that capture infrared light. These are commonly used for night vision surveillance and for thermal imaging. Thermal imaging cameras are already used in veterinary sciences as a diagnostic tool in clinical examination. The images can be used to identify differences in external/skin temperature that can be related to inflammatory process, infection, necrosis, and other symptoms. In research, infrared thermography (IRT) has been used to identify mastitis (Hovinen et al., 2008; Polat et al., 2010; Sathiyabarathi et al., 2016), and digital dermatitis (Alsaaod et al., 2014) in dairy cattle. However, the estimated local temperature can vary according to external factors, such as environmental temperature, wind speed, or other factors, such as operator and camera positioning (Byrne et al., 2017). Thus, applications for disease monitoring have been challenging, and are still generally based on semi-automated CVS.

1.4 Conclusion

Even though the idea of developing CVSs for automatic monitoring and measurement of traits of interest in live animals is not new, there are still several areas that demand intense research before efficient applications can be deployed to farmers. Research has corroborated that it is possible to use images to evaluate animal behavior, gait, body weight, and other traits in experimental conditions. Current challenges involve the development and implementation of reliable CVS for the autonomous acquisition of data from multiple traits in farm conditions.

1.5 References

Ahrendt, P., Gregersen, T., Karstoft, H., 2011. Development of a real-time computer vision system for tracking loose-housed pigs. Comput. Electron. Agric. 76, 169–174. https://doi.org/10.1016/j.compag.2011.01.011

Alsaaod, M., Syring, C., Dietrich, J., Doherr, M.G., Gujan, T., Steiner, A., 2014. A field trial of infrared thermography as a non-invasive diagnostic tool for early detection of digital dermatitis in dairy cows. Vet. J. 199, 281–285. https://doi.org/10.1016/j.tvjl.2013.11.028

Arganda-Carreras, I., Turaga, S.C., Berger, D.R., Cireşan, D., Giusti, A., Gambardella, L.M., Schmidhuber, J., Laptev, D., Dwivedi, S., Buhmann, J.M., Liu, T., Seyedhosseini, M., Tasdizen, T., Kamentsky, L., Burget, R., Uher, V., Tan, X., Sun, C., Pham, T.D., Bas, E., Uzunbas, M.G., Cardona, A., Schindelin, J., Seung, H.S., 2015. Crowdsourcing the creation of image segmentation algorithms for connectomics. Front. Neuroanat. 9, 142. https://doi.org/10.3389/fnana.2015.00142

Barnard, S., Calderara, S., Pistocchi, S., Cucchiara, R., Podaliri-Vulpiani, M., Messori, S., Ferri, N., 2016. Quick, Accurate, Smart: 3D Computer Vision Technology Helps Assessing Confined Animals' Behaviour. PLoS One 11, e0158748. https://doi.org/10.1371/journal.pone.0158748

Bay, H., Tuytelaars, T., Van Gool, L., 2006. SURF: Speeded Up Robust Features, in: European Conference on Computer Vision. Springer, Berlin, Heidelberg, pp. 404–417. https://doi.org/10.1007/11744023_32

Boyle, W.S., Smith, G.E., 1973. Buried channel charge coupled devices. US3792322A.

Burger, W., Burge, M.J., 2016. Digital image processing : An Algorithmic Introduction Using Java, 2nd ed. Springer London, London. https://doi.org/10.1007/978-1-4471-

6684-9
Byrne, D.T., Berry, D.P., Esmonde, H., McHugh, N., 2017. Temporal, spatial, inter-, and intra-cow repeatability of thermal imaging. J. Anim. Sci. 95, 970–979. https://doi.org/10.2527/jas2016.1005
Calonder, M., Lepetit, V., Strecha, C., Fua, P., 2010. BRIEF: Binary Robust Independent Elementary Features, in: European Conference on Computer Vision. Springer, Berlin, Heidelberg, pp. 778–792. https://doi.org/10.1007/978-3-642-15561-1_56
Cannell, R.C., Tatum, J.D., Belk, K.E., Wise, J.W., Clayton, R.P., Smith, G.C., 1999. Dual-component video image analysis system (VIASCAN) as a predictor of beef carcass red meat yield percentage and for augmenting application of USDA yield grades. J. Anim. Sci. 77, 2942. https://doi.org/10.2527/1999.77112942x
Cross, H.R., Gilliland, D.A., Durland, P.R., Seideman, S., 1983. Beef Carcass Evaluation by Use of a Video Image Analysis System. J. Anim. Sci. 57, 908–917. https://doi.org/10.2527/jas1983.574908x
Doeschl-Wilson, A.B., Green, D.M., Fisher, A. V, Carroll, S.M., Schofield, C.P., Whittemore, C.T., 2005. The relationship between body dimensions of living pigs and their carcass composition. Meat Sci. 70, 229–240. https://doi.org/10.1016/j.meatsci.2005.01.010
Folkestad, A., Wold, J.P., Rørvik, K.-A., Tschudi, J., Haugholt, K.H., Kolstad, K., Mørkøre, T., 2008. Rapid and non-invasive measurements of fat and pigment concentrations in live and slaughtered Atlantic salmon (Salmo salar L.). Aquaculture 280, 129–135. https://doi.org/10.1016/j.aquaculture.2008.04.037
Gomes, R.A., Monteiro, G.R., Assis, G.J.F., Busato, K.C., Ladeira, M.M., Chizzotti, M.L., 2016. Technical note: Estimating body weight and body composition of beef cattle trough digital image analysis. J. Anim. Sci. 94, 5414–5422. https://doi.org/10.2527/jas.2016-0797
Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press, Cambridge, MS.
Goodfellow, I.J., Warde-Farley, D., Mirza, M., Courville, A., Bengio, Y., 2013. Maxout Networks.
Gregory, R.L., 1978. Eye and Brain : The psychology of seeing, Third. ed. McGraw-Hill, New York.
Hough, P.V.C., 1962. METHOD AND MEANS FOR RECOGNIZING COMPLEX PATTERNS. US 3069654.
Hovinen, M., Siivonen, J., Taponen, S., Hänninen, L., Pastell, M., Aisla, A.-M., Pyörälä, S., 2008. Detection of Clinical Mastitis with the Help of a Thermal Camera. J. Dairy Sci. 91, 4592–4598. https://doi.org/10.3168/jds.2008-1218
Jabbar, K.A., Hansen, M.F., Smith, M.L., Smith, L.N., n.d. Locomotion Traits of Dairy Cows from Overhead Three-Dimensional Video.
Jackman, P., Sun, D.-W., Allen, P., 2011. Recent advances in the use of computer vision technology in the quality assessment of fresh meats. Trends Food Sci. Technol. 22, 185–197. https://doi.org/10.1016/j.tifs.2011.01.008
Jackman, P., Sun, D.-W., Allen, P., Brandon, K., White, A.-M., 2010. Correlation of consumer assessment of longissimus dorsi beef palatability with image colour, marbling and surface texture features. Meat Sci. 84, 564568. https://doi.org/10.1016/j.meatsci.2009.10.013

Kashiha, M.A., Bahr, C., Haredasht, S.A., Ott, S., Moons, C.P.H., Niewold, T.A., Ödberg, F.O., Berckmans, D., 2013a. The automatic monitoring of pigs water use by cameras. Comput. Electron. Agric. 90, 164–169. https://doi.org/10.1016/j.compag.2012.09.015

Kashiha, M.A., Bahr, C., Ott, S., Moons, C.P.H., Niewold, T.A., Ödberg, F.O., Berckmans, D., 2014a. Automatic weight estimation of individual pigs using image analysis. Comput. Electron. Agric. 107, 38–44. https://doi.org/10.1016/j.compag.2014.06.003

Kashiha, M.A., Bahr, C., Ott, S., Moons, C.P.H., Niewold, T.A., Ödberg, F.O., Berckmans, D., 2013b. Automatic identification of marked pigs in a pen using image pattern recognition. Comput. Electron. Agric. 93, 111–120. https://doi.org/10.1016/j.compag.2013.01.013

Kashiha, M.A., Bahr, C., Ott, S., Moons, C.P.H., Niewold, T.A., Tuyttens, F., Berckmans, D., 2014b. Automatic monitoring of pig locomotion using image analysis. Livest. Sci. 159, 141–148. https://doi.org/10.1016/j.livsci.2013.11.007

Kashiha, M.A., Bahr, C., Ott, S., Moons, C.P.H., Niewold, T.A., Tuyttens, F., Berckmans, D., 2013c. Automatic Monitoring of Pig Activity Using Image Analysis. Springer, Cham, pp. 555–563. https://doi.org/10.1007/978-3-319-02895-8_50

Kheradmand, A., Milanfar, P., 2015. Non-Linear Structure-Aware Image Sharpening with Difference of Smoothing Operators. Front. ICT 2, 22. https://doi.org/10.3389/fict.2015.00022

Kimiya, T., Sivertsen, A.H., Heia, K., 2013. VIS/NIR spectroscopy for non-destructive freshness assessment of Atlantic salmon (Salmo salar L.) fillets. J. Food Eng. 116, 758–764. https://doi.org/10.1016/j.jfoodeng.2013.01.008

Kongsro, J., 2014. Estimation of pig weight using a Microsoft Kinect prototype imaging system. Comput. Electron. Agric. 109, 32–35. https://doi.org/10.1016/j.compag.2014.08.008

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet Classification with Deep Convolutional Neural Networks, in: Neural Information Processing Systems. pp. 1097–1105.

Lao, F., Brown-Brandl, T., Stinn, J.P., Liu, K., Teng, G., Xin, H., 2016. Automatic recognition of lactating sow behaviors through depth image processing. Comput. Electron. Agric. 125, 56–62. https://doi.org/10.1016/j.compag.2016.04.026

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539

Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86, 2278–2324. https://doi.org/10.1109/5.726791

Leutenegger, S., Chli, M., Siegwart, R.Y., 2011. BRISK: Binary Robust invariant scalable keypoints, in: International Conference on Computer Vision. IEEE, pp. 2548–2555. https://doi.org/10.1109/ICCV.2011.6126542

Lin, S.S.P., 1978. Automated pattern recognition of beef. Kansas State University.

Lind, N.M., Vinther, M., Hemmingsen, R.P., Hansen, A.K., 2005. Validation of a digital video tracking system for recording pig locomotor behaviour. J. Neurosci. Methods 143, 123–132. https://doi.org/10.1016/j.jneumeth.2004.09.019

Lister, M., 1995. The photographic image in digital culture. Routledge, London.

Litjens, G., Sánchez, C.I., Timofeeva, N., Hermsen, M., Nagtegaal, I., Kovacs, I., Hulsbergen - van de Kaa, C., Bult, P., van Ginneken, B., van der Laak, J., 2016.

Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. Sci. Rep. 6, 26286. https://doi.org/10.1038/srep26286

Mattachini, G., Riva, E., Bisaglia, C., Pompe, J.C.A.M., Provolo, G., 2013. Methodology for quantifying the behavioral activity of dairy cows in freestall barns. J. Anim. Sci. 91, 4899–4907. https://doi.org/10.2527/jas.2012-5554

Mendes, E.D.M., Carstens, G.E., Tedeschi, L.O., Pinchak, W.E., Friend, T.H., 2011. Validation of a system for monitoring feeding behavior in beef cattle. J. Anim. Sci. 89, 2904–2910. https://doi.org/10.2527/jas.2010-3489

Nakagomi, K., Shimizu, A., Kobatake, H., Yakami, M., Fujimoto, K., Togashi, K., 2013. Multi-shape graph cuts with neighbor prior constraints and its application to lung segmentation from a chest CT volume. Med. Image Anal. 17, 62–77. https://doi.org/10.1016/j.media.2012.08.002

Nasirahmadi, A., Edwards, S.A., Matheson, S.M., Sturm, B., 2017. Using automated image analysis in pig behavioural research: Assessment of the influence of enrichment substrate provision on lying behaviour. Appl. Anim. Behav. Sci. https://doi.org/https://doi.org/10.1016/j.applanim.2017.06.015

Nunes, J., Piquerez, M., Pujadas, L., Armstrong, E., Fernández, A., Lecumberry, F., 2015. Beef quality parameters estimation using ultrasound and color images. BMC Bioinformatics 16 Suppl 4, S6. https://doi.org/10.1186/1471-2105-16-S4-S6

Oczak, M., Viazzi, S., Ismayilova, G., Sonoda, L.T., Roulston, N., Fels, M., Bahr, C., Hartung, J., Guarino, M., Berckmans, D., Vranken, E., 2014. Classification of aggressive behaviour in pigs by activity index and multilayer feed forward neural network. Biosyst. Eng. 119, 89–97. https://doi.org/10.1016/j.biosystemseng.2014.01.005

Otsu, N., 1979. A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man. Cybern. 9, 62–66.

Ott, S., Moons, C.P.H., Kashiha, M.A., Bahr, C., Tuyttens, F.A.M., Berckmans, D., Niewold, T.A., 2014. Automated video analysis of pig activity at pen level highly correlates to human observations of behavioural activities. Livest. Sci. 160, 132–137. https://doi.org/10.1016/j.livsci.2013.12.011

Polat, B., Colak, A., Cengiz, M., Yanmaz, L.E., Oral, H., Bastan, A., Kaya, S., Hayirli, A., 2010. Sensitivity and specificity of infrared thermography in detection of subclinical mastitis in dairy cows. J. Dairy Sci. 93, 3525–3532. https://doi.org/10.3168/jds.2009-2807

Polesel, A., Ramponi, G., Mathews, V.J., 2000. Image enhancement via adaptive unsharp masking. IEEE Trans. Image Process. 9, 505–510. https://doi.org/10.1109/83.826787

Pratt, W.K., 2007. Digital image processing, Fourth. ed. John Wiley & Sons, Inc., Hoboken, New Jersey.

Chapter 2: Deep Learning image segmentation for extraction of fish body measurements and prediction of body weight and carcass traits in Nile tilapia

2.1 Abstract

Individual measurement of traits of interest is of great importance in aquaculture, both for production systems and for breeding programs, and also for the assessment of animal resources. Most of the current methods are based on manual measurements, which are laborious and stressful to the animals. Therefore, the development of fast, precise and indirect measurement methods for traits such as body and carcass weights, as well as biometric traits, such as length and height, are of great interest. An appealing way to take noninvasive measurements in such a fashion is through computer vision. Hence, the objectives in the current work were to: 1) devise a computer vision system (CVS) for autonomous measurement of Nile tilapia body area, length, height, and eccentricity, and 2) develop linear models for prediction of fish body weight, carcass weight, and carcass yield. For the first step in the development of the CVS, the crowdsourcing service Mechanical Turk was used for labeling image pixels into three classes: background, fish fins, and body areas. This labeled dataset was then used for training of Deep Learning Networks for automatic segmentation of the images into those three pixel categories. In a subsequent step, the segmentation results obtained from the best network were used for extraction of fish body area, length, height, and eccentricity information. These variables were finally used as covariates in linear models for prediction of body weight, carcass

weight, and carcass yield. The networks were evaluated regarding their accuracy to predict the classes of pixels in the test set. A network with an input image of 0.2 times the original size and four encoder/decoder layers achieved the best results for intersection over union on the test set of 99, 90, and 64 percent for background, fish body and fin areas, respectively. For the prediction of body and carcass weights, the linear model including only body area as a predictor variable achieved a predictive R^2 of 0.95 and 0.94 for fish body and carcass weights, respectively. However, for prediction of carcass yield, any model evaluated presented desirable predictions. Overall, the devised CVS was able to correctly differentiate fish body from background and fins, and the extracted area of the fish body could be successfully used for prediction of body and carcass weights.

2.2 Introduction

Fish consumption is growing vertiginously worldwide and, since capture is stagnated with most of the fisheries resources already overexploited, this expansion in consumption is directly related to an increase in aquaculture production. Specifically, aquaculture has grown from 47 million tons in 2006 to 80 million tons in 2016 (FAO, 2018), and thus, is an important protein source for the growing human population. Nevertheless, to continue with such a growth rate, the industry and the academic community need to constantly search for ways to increase productivity while controlling the production environmental impact.

An important tool to increase general aquaculture productivity and to improve traits of interest is genetic selection and breeding strategies. There are several known Nile tilapia breeding programs worldwide, and the main focus of those programs is mostly on

final body weight, although the ultimate interest is on meat production and quality (Gjedrem et al., 2012). However, direct measurement of meat traits, such as carcass yield and meat quality is impossible on live animals. Possible solutions are via selection based on the performance of relatives or via indirect selection using indicator traits. The selection based on data from relatives has the drawback of reduced selection intensity since a portion of the fish need to be harvested for data collection. For selection based on indicator traits, its success depends directly on the genetic correlation between the indicator traits and the traits of interest. Previous studies shown that body and carcass weights are highly correlated with biometric traits, such as height, length, and width (Fernandes et al., 2015; Turra et al., 2018). However, direct measurement of these traits can be time-consuming and cumbersome. Moreover, it can be stressful for the fish, negatively affecting their growth performance and even resulting in increased mortality.

In this context, fast and non-invasive alternatives to measure traits of interest would be of great help to farmers and biologists. Computer vision systems are interesting in this regard since pictures or videos from each animal can be acquired and used to measure several traits remotely. In addition, recent advances in artificial intelligence are showing promising results in areas such as image processing, segmentation, and prediction of characteristics of interest (Goodfellow et al., 2016). However, in the available literature, there are only a few attempts of using computer vision systems (CVS) in aquaculture. A recent attempt is the IMAFISH (Navarro et al., 2016), which was developed for three marine fish species. Even though the approach presented satisfactory results in terms of the correlation between the image and manual measurement of biometric traits, it requires the fish to be already frozen and the fins to be trimmed. Therefore, our objective was to

develop an automated CVS for measurement of biometric traits of interest and prediction on body weight (BW), carcass weight (CW) and yield (CY) of live Nile tilapia.

2.3 Material and Methods

2.1. Fish Breeding and Grow out

A total of 1,653 Nile tilapia were used in this study. The fish were produced by mating 43 males and 86 females selected from the broodstock at the Aquaculture Laboratory (LAQUA) from the Veterinary School of the Federal University of Minas Gerais (UFMG), Brazil. The experiment was conducted following rigorous animal handling procedures that are in compliance with federal and institutional regulations regarding proper animal care practices (CONCEA, 2016). The breeding occurred for 58 days, from February to April 2014. During the breeding season, males were placed with 4 females in 1 m^3 tanks for reproduction. After one week of reproduction, females with eggs in their mouth were carefully restrained and had the eggs removed and sent to incubators. The males stayed in the reproduction tanks until they mated with 2 females, thus producing 2 full-sib families. Each full-family stayed in the incubators for approximately one week; after that, the fish from the same family were reared in net tanks (*hapas*) until tagging. At approximately 60 days post-hatching, fish from the families closest in age were individually weighed, tagged and grouped for the grow-out phase. For tagging the animals, a passive integrated transponder (PIT) tag was inserted intramuscular close to the dorsal fin. The fish had an average initial BW of 19.0 (± 11.8) g. The grow-out phase had a duration of approximately 220 days (min of 140 and max of 300 days). Thus, at harvest, the fish had an average age of 280 (± 36) days of life. Prior to harvest, the fish were again individually

weighed, after having feed withdrawn to allow their digestive system to empty its content. In this experiment, the CW was considered as the weight of the fish body without scales and viscera, and CY the ratio between CW and BW. The fish had average final BW of 523.8 (± 224.4) g, CW of 294.3 (± 126) g, and CY of 56.25 (± 2.89) %.

2.2 Image Data Acquisition and Data Sets

At the end of the grow-out phase, along with the measurement of BW and carcass traits, pictures were taken of each fish using a Sony DSCWX220 (Sony, Tokyo - JP) digital camera equipped with complementary metal–oxide–semiconductor (CMOS) sensor with a resolution of 18.2 megapixels. The camera was configured for maximum resolution, fast picture mode, and adaptive focus. For stability and improvement of the image acquisition, the camera was fixed via a table-top stand mount that was coupled to a table prepared to hold the fish. The table was covered with a green background for maximum contrast with

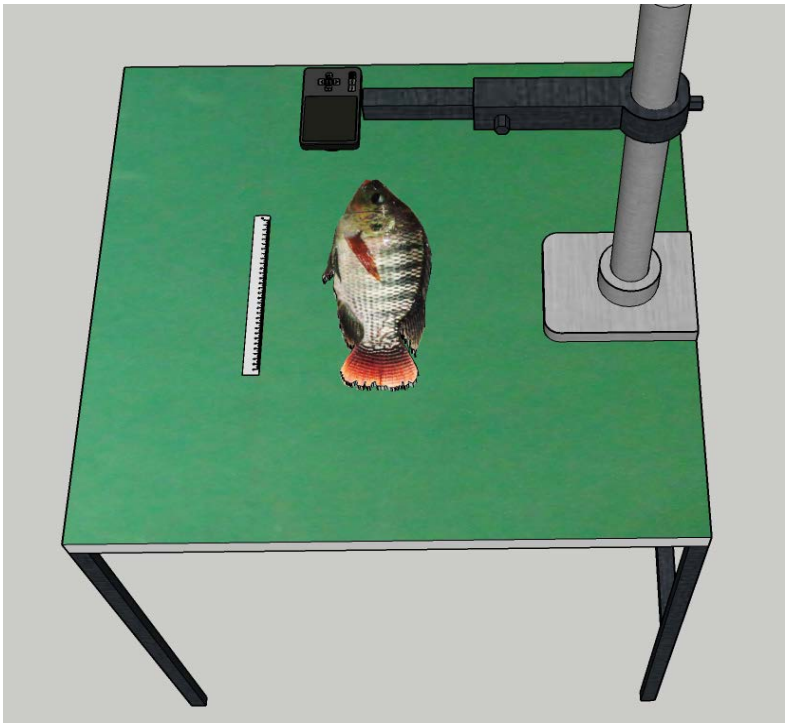

Figure 2.1 *Example of the station set up for image acquisition with a green background and a digital camera positioned on top of fish at an average distance of 0.5 m.*

the fish. The table also had rulers fixed on top of the green background for the conversion of pixel measures to millimeters (Figure 2.1). At the moment the pictures were taken, the fish had their identification PIT tag read for tracking their identification, and their body weight was measured using an electronic scale.

The dataset was split in two parts, dataset 1 had the 822 fish produced in the first month, whereas dataset 2 had the 831 fish produced in the second month of reproduction. The images from the fish from dataset 1 were used for training and evaluation of the image segmentation models for pixel classification. Thus, these images had a label matrix where each pixel had a final class that could be one of the previously defined classes, i.e. background, fish body or fins. For evaluation of the image segmentation models, the dataset 1 was split into training and testing sets. The training set contained 60% of the images (493 images) and the test set, 40% (329 images). Fish in dataset 2 did not have labeled images and were then used for the validation of the final prediction models for BW, CW, and CY. As a final assessment, the developed segmentation algorithm was tested on 40 images of other species of fish publicly available on the web. These images were either from more related fish species, such as the blue (*Oreochromis aureus* - 7 images), mossambicus (*Oreochromis mossambicus* - 4 images), red (*O. aureus x O. mossambicus* - 3 images), zanzibar (*Oreochromis urolepis* - 3 images) and zili (*Coptodon zillii* - 3 images) tilapias, or images from carp species, such as common (*Cyprinus carpio* - 11 images), grass (*Ctenopharyngodon idella* - 7 images) and big head (*Hypophthalmichthys nobilis* - 4 images) carps.

2.3 Image Segmentation and Analysis

There were two different image segmentation algorithms used in the present work. The first corresponded to a color threshold algorithm for removal of the background, and identification and segmentation of the whole fish. The algorithm was written and implemented in Matlab R2017b (The MathWorks, 2017), and comprised the following steps: 1) downsize of each image to 1944 x 2952 pixels; 2) conversion of images from RGB to CIE 1976 L*a*b space; 3) removal of green background via Global Image Threshold on the image channels based on their histogram distribution; and 4) Removal of noise on the resultant black and white images by filling holes in the fish mask and opening the image with a disk kernel with a diameter of 50 pixels (Figure 2.2).

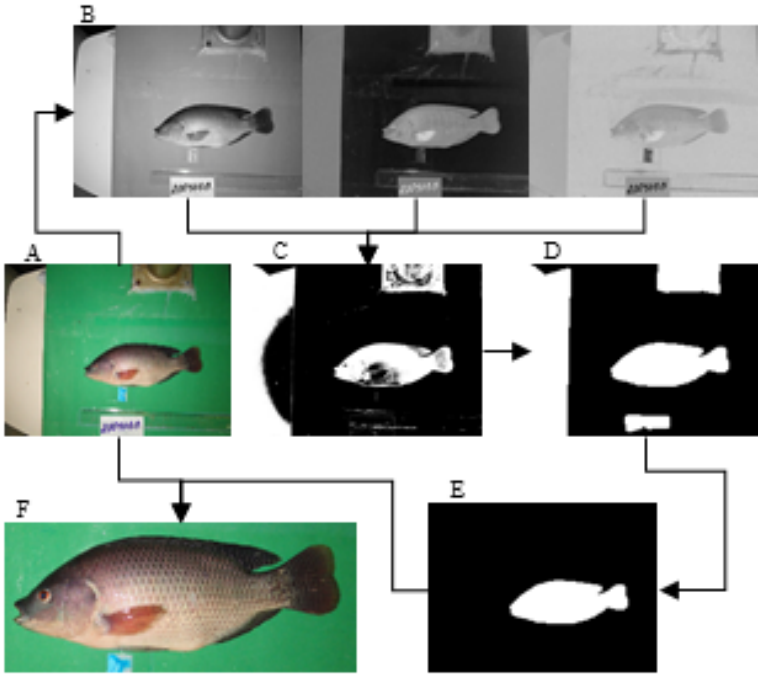

Figure 2.2 *Representation of background removal and fish identification. A) Original image. B) The channels of the image converted to the CIE 1976 L*a*b color space. C) Global thresholding on the LAB channels. D) Morphological image opening and filling. E) The binary mask of the identified fish. F) Fish cropped from the original image using the binary mask.*

The algorithm described above was used to generate cropped images containing only the fish. Then, each cropped image was used in the following steps for labeling of fish body area (BA) and fins. The fish BA in each image was manually labeled using a web application developed in house. The manual labels of fish BA were then grouped with the resultant segmented images from the color threshold method for the construction of a labeled dataset where each pixel in the original image was labeled as one of the following classes: background, fish body, or fish fins (Figure 2.3). To generate this dataset in a fast and robust manner, the manual image labeling was crowdsourced in Amazon Mechanical Turk (MTurk) (Amazon, Seattle-WA). In the crowdsourcing step, each image was labeled by an anonymous worker, and all labels were later curated and approved by a trained evaluator. If a label was not accepted, the same image was sent back to the crowdsourcing service where it was assigned to a different worker.

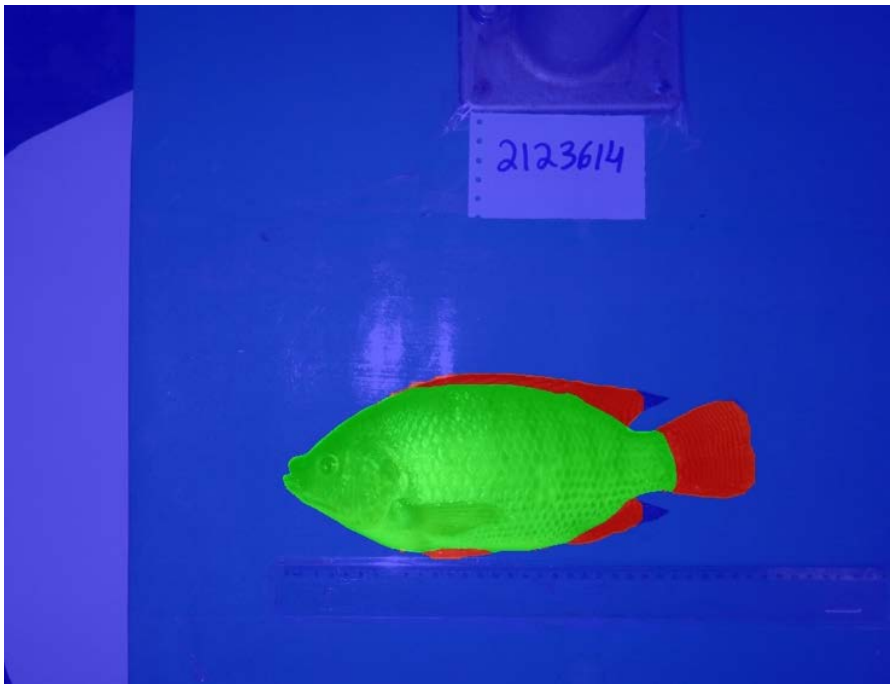

Figure 2.3 *Example of the labeled image in the crowd-sourcing step, with pixel labels classes specified as background (blue), fish body (green) and fish fins (red).*

The models used for image segmentation are known as segmentation networks, a class of Deep Learning models for which the main goal is to classify every input node properly. More specifically, the image segmentation models were based on the encoder-decoder network known as SegNet (Badrinarayanan et al., 2015). The segmentation networks evaluated had encoder layers that were composed of a convolutional layer (CV1), batch normalization (BN1), rectified linear unit (ReLu1), convolutional layer (CV2), batch normalization (BN2), rectified linear unit (ReLu2), and a max pooling layer (MP). Therefore, each encoder had a total of 7 layers. For each encoder in the model, there was a respective decoder that had a max un-pooling layer (MU), a convolutional layer (CV2), batch normalization (BN2), rectified linear unit (ReLu2), convolutional layer (CV1), batch normalization (BN1), and rectified linear unit (ReLu1). Finally, a fully connected layer and a final pixel-wise classification layer were added to the networks, coupled to the last decoder layer. To improve training performance, the training set was augmented using image rotation, reflection, and translation.

The networks evaluated diverged in the size of the input image, which could be 0.1, 0.2, 0.3 and 0.4 times of the original size (1944×2592 pixels), and the number of complex encoder-decoder layers, which ranged from 1 to 5. For evaluation of predictive quality of pixel classification, the overall prediction accuracy and the intersection over union (IoU) ratio were calculated. The IoU is defined by, $IoU_{A,B} = (A \cap B)/(A \cup B)$, where in our case A and B represent the predicted and expected areas. Also, the networks were compared on the usage of computational resources, such as disk space for model allocation (MB), computational time (s) and random allocation memory (RAM) (MB) required for evaluation of new frames.

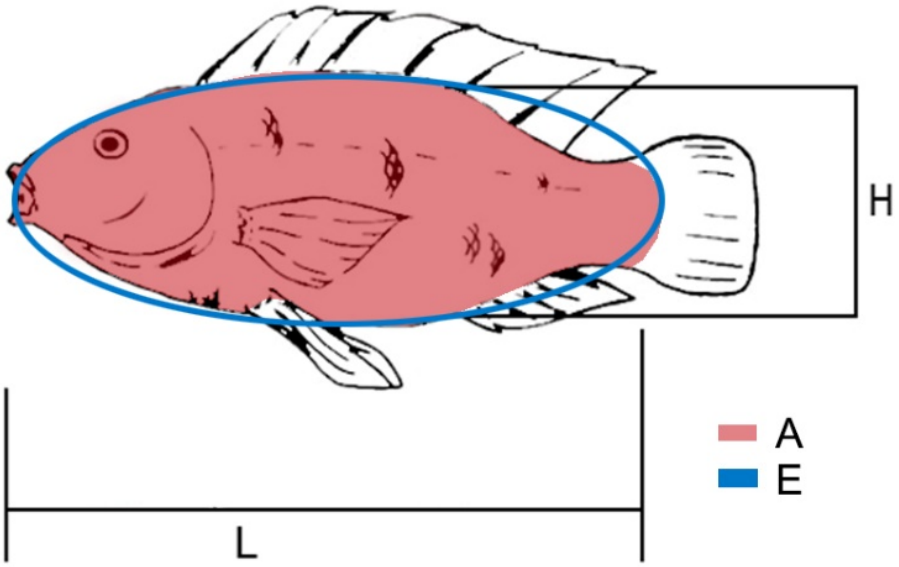

Figure 2.4 *Extracted body measurements from image: body area (A), height (H), length (L) and ellipsis (E) that has the same second moments of the body area.*

From the segmented images, the value of each pixel in the images was linearly transformed to centimeters by the measurement of the correspondent distance in pixels from the ends of the ruler affixed in the measurement table. Fish body area (BA), length (L), and height (H) in centimeters were then calculated using this linear transformation. The eccentricity (E) of the fish body was calculated as the ratio between the foci and the major axis length of the ellipsis with the same major and minor axis as the fish body area (Figure 2.4).

2.4. Statistical Analysis

Estimates of *Pearson's* correlation between body measurements and body weight (BW), carcass weight (CW) and, carcass yield (CY) were computed. Multiple linear regression models were evaluated for prediction of BW, CW, and CY. These models differed in terms of the inclusion of specific combinations of the body measurements and possible interactions. The search for the best possible model on the training set was

performed using the Bayesian Information Criteria (BIC) to rank the models and a genetic algorithm search (Calcagno and Mazancourt, 2010). In this analysis, the fish body predicted from the best segmentation model (model with depth 4 and input image size of 20%) developed in the previous section was used as input for automated acquisition of fish body measurements. For this purpose, the segmentation model was retrained using the dataset 1, and the predicted segmented fish body from both datasets was used. The predictive models were fitted using only the fish that had information on BW, CW and CY from dataset 1 (that was 764 fish labeled images) and evaluated on dataset 2 (that was 765 fish). These statistical analyses were conducted in R (R Core Team, 2016) using the packages MASS, *glmulti* (Calcagno and Mazancourt, 2010), and ggplot2 (Wickham, 2009).

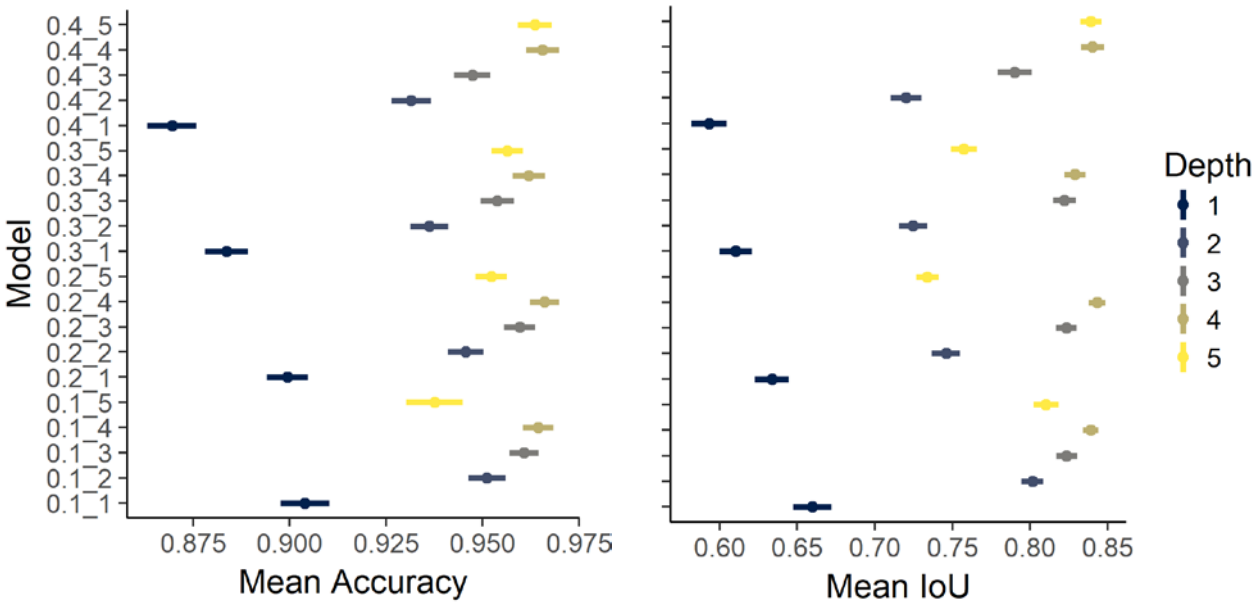

Figure 2.5 *Average model accuracy and intersection over union (IoU) and respective 95% CI for an input image with a size that is of 0.1 to 0.4 times of the original size and encoder depth varying from 1 to 5 stacks of layers on the test dataset.*

2.4 Results

2.4.1 Semantic Segmentation

Figure 2.5 presents the mean accuracy and intercept over union (IoU) and their respective 95% confidence interval (CI) for the segmentation networks evaluated. The results show an increase in both overall accuracy and IoU as the depth of the network increased, reaching a maximum at depth 4. Similarly, an increase in IoU was also observed for each of the three different pixel classes when evaluated separately, as the depth of the model increased (Figure 2.6). However, the background and body had higher IoU with the best models achieving 0.99 and 0.90, respectively, while for fins the best models achieved at maximum 0.65.

The differences in classification performance between fish fin versus body or background were expected. This is mostly because the fin class had the lowest number of pixels and it was also very difficult to differentiate the boundaries between fish fins and body or in some cases even with the background, even for human evaluators. Figure 2.7

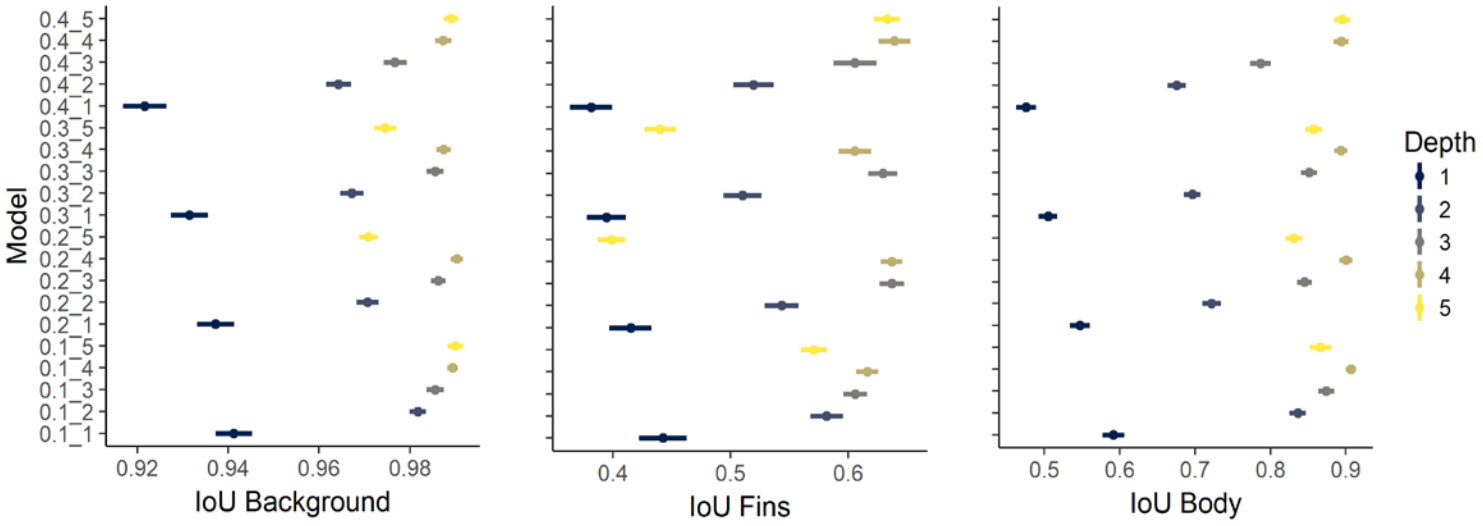

Figure 2.6 *Average intersection over union (IoU) for the 3 classes (Background, Fins, and Body), and respective 95% CI for an input image of 0.1 to 0.4 times of the original size and encoder depth from 1 to 5 stacks of layers on the test dataset.*

presents examples of the labeled images with their correspondent pixel classification image. By visual inspection of the resulting pixel classifications versus the input label image it is possible to observe that background pixels near to the fin or on the border of the fish were more prone to be erroneously classified as fin (Figures 2.7.2 and 2.7.3). Also, patches of the background that had more texture due to water splashes or higher light reflection could be incorrectly classified as fish (Figures 2.7.2, 2.7.3 and 2.7.4).

Regarding the images of other fish species retrieved from the web, the results were, in general, worse than what was observed for the Nile tilapia dataset, with an average accuracy of 0.57 and IoU of 0.52, 0.34 and 0.1 for background, body, and fins respectively (Table 2.1). This result is reflected in the visual inspection of the images (Figure 2.8). It can be observed that images with fish that had either body shape or color different from the fish used to generate the models and/or with a more complex background, such as grass or a fishnet, presented the worse predictions.

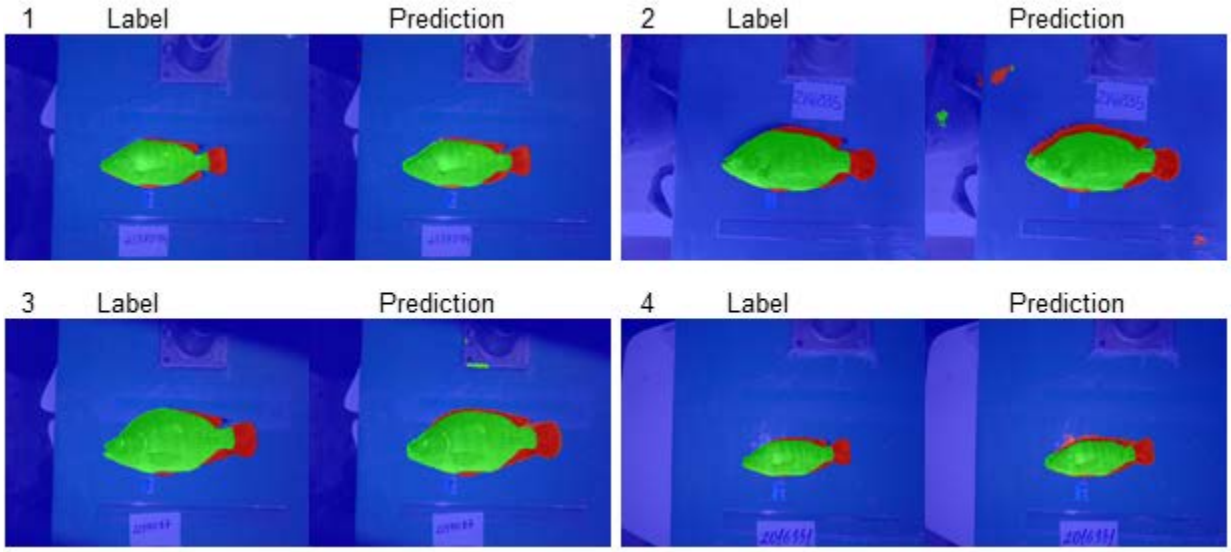

Figure 2.7 *Randomly selected examples of fish images from the test dataset showing the expected result (Label) and model output predicted pixel classification (Prediction) with the background, fish body and fins coded as blue, green and red respectively.*

Table 2.1 *Average and 95% confidence interval (CI) for accuracy (Acc) and intersection over union (IoU) for the best segmentation network on 40 fish images retrieved from the web*

	Acc	Mean IoU	IoU Background	IoU Fins	IoU Body
Average	0.573	0.318	0.526	0.100	0.340
95% CI	±0.068	± 0.049	±0.083	±0.031	±0.063

In the evaluation of computational resources used, there was an overall increase in disk and RAM usage along with computational time as the model complexity increased (Table 2.2). The main factor in increasing the demand for computational resources was the size of the input image. Also, the evaluation using graphics processing unit (GPU) was overall 20 to 40 times faster but used from 1.5 to 10 times more RAM than with central processing unit (CPU). It is also worth noting that the usage of RAM increased at a steeper rate on CPU evaluations compared to GPU. Thus, RAM usage on a CPU-based version

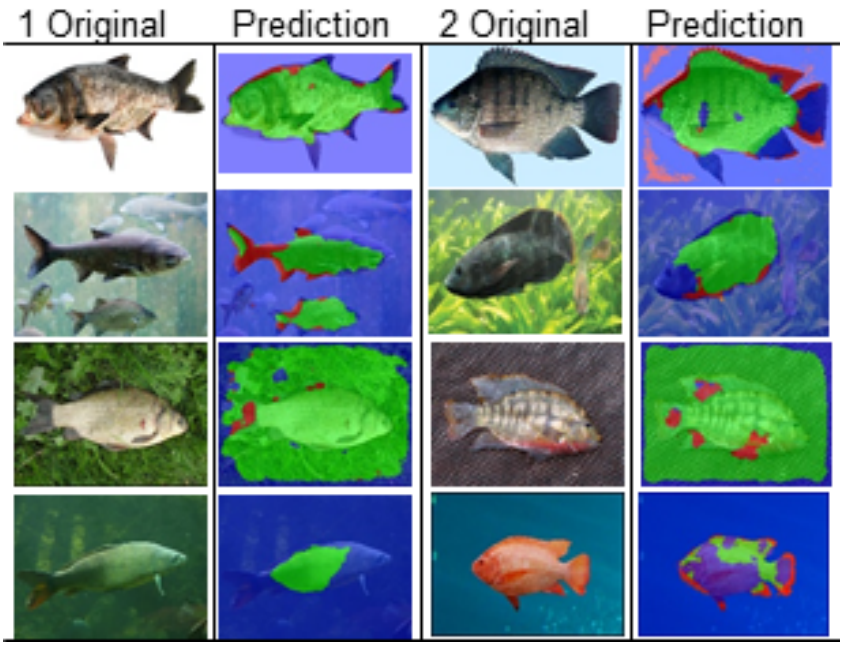

Figure 2.8 *Randomly selected examples of carp (1) and tilapia (2) images retrieved from the web showing the original image and model pixel classification (Prediction) with the background, fish body and fins coded as blue, green and red respectively.*

of the model probably would surpass the RAM usage of the GPU version for more complex models.

2.4.2 Image Derived Body Measurements and Body Weight Prediction

It was observed that the pixel length ranged from 0.021 to 0.026 cm, due to the difference in the height that the camera was attached in each day of measurement. However, there was no difference in the pixel length at the plane of the table for fish measured on the same day.

Table 2.3 shows the estimated *Pearson's* correlation between body measurements, BW, CW, and CY for both datasets. Correlations between body measurements and BW or CW were positive and high, ranging from 0.87 for H and CW in dataset 2 to 0.96 for BA and BW in dataset 1. However, E presented moderate to low negative correlations with the other variables in dataset 1, while in dataset 2 the correlations were mostly low and close to zero. Estimates of correlations between CY and the other variables were mostly low and positive.

Table 2.2 *Comparison of the Deep Learning architectures evaluated on image size (pixels), number (N) of encoders, layers and weights, computational resources used as disk space (MB), computational time (s) and random allocation memory (RAM) required on evaluation step to predict pixels class in GPU or CPU*

Image Size	N Encoders	N Layers	N Weights	N Nodes	Model Size (MB)	GPU[1] RAM (MB)	GPU Time (s)	CPU[2] RAM (MB)	CPU Time (s)
200 x 237	1	17	77967	398364	1.8	1202.46	0.06	91	1.1
	2	31	226447	546844	2.84	1511.664	0.07	99	1.66
	3	45	374927	695324	3.89	1322.706	0.09	105	1.81
	4	59	523407	843804	4.93	1288.35	0.1	117	1.82
	5	73	671887	992284	5.97	1202.46	0.11	125	1.82
390 x 520	1	17	77967	1294764	4.4	1717.8	0.15	270	4.35
	2	31	226447	834844	5.08	1889.58	0.19	357	6.3
	3	45	374927	983324	6.13	1803.69	0.2	405	6.8
	4	59	523407	1131804	7.17	1803.69	0.21	465	6.9
	5	73	671887	1280284	8.25	1631.91	0.23	461	6.9
585 x 780	1	17	77967	1446864	7.67	1975.47	0.3	750	10.7
	2	31	226447	1595344	8.71	1975.47	0.38	1100	15.75
	3	45	374927	1743824	9.75	2061.36	0.39	1240	16.5
	4	59	523407	1892304	10.8	2147.25	0.4	1230	16.85
	5	73	671887	2040784	11.8	2147.25	0.42	1300	17
780 x 1040	1	17	77967	2511564	12.5	2576.7	0.57	1640	24.5
	2	31	226447	2660044	10.9	3006.15	0.79	2207	31
	3	45	374927	2808524	12	3006.15	0.8	2268	33.5
	4	59	523407	2957004	13	3092.04	0.82	2245	33.7
	5	73	671887	3105484	14.1	3092.04	0.84	2272	33.7

[1] The graphic processing unit (GPU) used in this study was an NVIDIA Quadro M4000. [2] The central processing unit (CPU) used in the current study was an Intel Core i7-7700K.

Table 2.3 *Pearson's correlation between fish body area (BA), length (L), height (H), eccentricity (E), body weight (BW), carcass weight (CW) and carcass yield (CY) on dataset 1 (below diagonal) and on dataset 2 (above diagonal)*

	BA	**L**	**H**	**E**	**BW**	**CW**	**CY**
BA	-	$0.97^{(0.00)}$	$0.97^{(0.00)}$	$-0.40^{(0.00)}$	$0.98^{(0.00)}$	$0.97^{(0.00)}$	$0.08^{(0.02)}$
L	$0.97^{(0.00)}$	-	$0.92^{(0.00)}$	$-0.20^{(0.00)}$	$0.93^{(0.00)}$	$0.93^{(0.00)}$	$0.06^{(0.11)}$
H	$0.98^{(0.00)}$	$0.92^{(0.00)}$	-	$-0.55^{(0.00)}$	$0.95^{(0.00)}$	$0.94^{(0.00)}$	$0.07^{(0.04)}$
E	$-0.56^{(0.00)}$	$-0.37^{(0.00)}$	$-0.69^{(0.00)}$	-	$-0.40^{(0.00)}$	$-0.40^{(0.00)}$	$-0.06^{(0.11)}$
BW	$0.97^{(0.00)}$	$0.92^{(0.00)}$	$0.96^{(0.00)}$	$-0.63^{(0.00)}$	-	$0.99^{(0.00)}$	$0.07^{(0.06)}$
CW	$0.96^{(0.00)}$	$0.90^{(0.00)}$	$0.95^{(0.00)}$	$-0.63^{(0.00)}$	$0.99^{(0.00)}$	-	$0.18^{(0.00)}$
CY	$0.21^{(0.00)}$	$0.18^{(0.00)}$	$0.22^{(0.00)}$	$-0.21^{(0.00)}$	$0.22^{(0.00)}$	$0.34^{(0.00)}$	-

Superscripts are the respective p-values.

Regarding the predictions of BW and CW, it was observed that the simpler model that included only BA as predictor variable had better performance on the testing dataset than more complex models (Table 2.4). This model achieved, for BW and CW, predicted square correlations (R^2) of 0.92 and 0.90 and mean absolute errors (MAE) of 63.62 g and 40.21 g, respectively. Since the average BW was 540 g and CW was 300 g, the MAE was on the order of 11.8 and 13.4% for BW and CW, respectively. On the other hand, the MAE for the model including only BA as predictor variable was of 2.3 for prediction of CY, which corresponds to an error of about 4%, since the average CY was of 56%. However, the predicted R^2 was lower than 0.22 for every predictive model evaluated for CY.

Table 2.4 *Comparison of selected linear models for predictions of body weight (BW), carcass weight (CW) and carcass yield (CY) accounting for area (BA), length (L), height (H) and eccentricity (E)*

		Training Dataset				Test Dataset			
Target Variable	**Input Variable**	**R^2**	**MAE[a]**	**RMSE[b]**	**SE[c]**	**R^2 [d]**	**MAE**	**RMSE**	**SE**
BW	**BA**	**0.95**	**35.47**	**48.32**	**1.74**	**0.95**	**62.33**	**77.45**	**2.80**
	L	0.84	59.66	83.49	3.01	0.87	79.17	101.78	3.68
	H	0.93	42.33	55.10	1.99	0.90	71.79	95.78	3.46
	BA+E	0.96	31.87	43.39	1.57	0.94	63.34	81.88	2.96
	L+E	0.94	39.28	53.00	1.91	0.91	74.22	93.87	3.39
	H+E	0.93	40.75	53.86	1.95	0.91	69.88	91.69	3.31
	BA+L+H+E	0.96	30.84	42.77	1.54	0.95	62.03	81.01	2.93
	BA + E×L	0.96	30.87	42.89	1.55	0.95	60.33	79.78	2.88
	BA+H+H×L+E×BA	0.96	28.68	39.66	1.43	0.95	59.62	81.98	2.96
CW	**BA**	**0.92**	**25.36**	**34.58**	**1.25**	**0.94**	**39.44**	**49.43**	**1.79**
	L	0.81	38.81	54.31	1.96	0.86	48.67	62.42	2.26
	H	0.91	27.82	37.44	1.35	0.88	44.94	59.76	2.16
	BA+E	0.94	22.97	31.36	1.13	0.93	40.23	52.85	1.91
	L+E	0.91	27.04	36.76	1.33	0.89	46.74	59.94	2.17
	H+E	0.91	27.42	37.17	1.34	0.89	44.16	58.13	1.88
	BA+L+H+E	0.94	22.24	30.67	1.11	0.93	39.18	52.13	1.88
	BA + E×L	0.94	22.33	30.73	1.11	0.93	38.17	51.39	1.86
	BA+H+H×L+E×BA	0.95	20.57	28.75	1.04	0.94	37.96	53.63	1.94
CY	BA	0.04	2.21	2.78	0.10	0.01	2.27	2.84	0.10
	L	0.03	2.22	2.80	0.10	0.00	2.28	2.84	0.10
	H	0.05	2.21	2.78	0.10	0.00	2.28	2.85	0.10
	BA+E	0.06	2.19	2.77	0.10	0.01	2.28	2.85	0.10
	L+E	0.05	2.20	2.77	0.10	0.00	2.29	2.86	0.10
	H+E	0.05	2.20	2.77	0.10	0.00	2.29	2.85	0.10
	BA+L+H+E	0.06	2.20	2.76	0.10	0.01	2.28	2.85	0.10
	BA+E×L	0.06	2.20	2.76	0.10	0.01	2.28	2.84	0.10
	BA+H+H×L+E×BA	0.06	2.19	2.76	0.10	0.01	2.29	2.86	0.10

[a] Mean Absolute Error; [b] Residual Mean Square Error; [c] Residual Standard Error; [d] Squared Correlation.

2.5 Discussion

In the current paper, a CVS for biometric measurement of fish bodies and prediction of fish BW, CW, and CY was presented. To the best of the authors' knowledge,

this is the first work to present such results for Nile tilapia. However, there are previous attempts in the literature to develop automated CVS for other fish species, such as seabass (*Dicentrarchus labrax, L.*) (Costa et al., 2013), olive flounder (*Paralichthys olivaceus*) (Jeong et al., 2013), gilthead seabream (*Sparus aurata L.*), meagre (*Argyrosomus regius*) and red porgy (*Pagrus pagrus*) (Navarro et al., 2016). These studies focused on the development of an automated CVS for indirect measurement and prediction of traits of interest. However, none of the methods presented was developed for live animals. Moreover, the images were taken in an environment with controlled illumination conditions, and for some species (e.g., seabass, meagre, and porgy) the fins were manually cut to improve predictions. Thus, the CVS presented in this manuscript is an improvement to these previous methods because the images of live fish were taken on multiple sites without the aid of controlled illumination.

A drawback of the proposed CVS is that it still requires the removal of fish from the water to set them on a table covered with a green background for the acquisition of the images. Moreover, as expected, the models developed are applicable only for Nile tilapia on a flat surface, as it became evident after the evaluation of the semantic segmentation model on images from other fish species and in different background conditions. For the 40 images of other fish species, the average accuracy dropped to 0.57 and the IoU to 0.32 (Table 2.1). By visual inspection of such images (Figure 2.8) it is possible to evaluate that the best results were achieved if the fish were on a more homogeneous background compared to texturized backgrounds, such as grass or inside a net. It is also worth noticing that the model performs better for fish with body color that is more dark brown,

similar to the Nile tilapia, since the best results were for the top images of Figure 2.8 representing a blue tilapia while the red tilapia on the bottom image was not identified.

2.5.1 Semantic Segmentation

From the several existing image classification and segmentation methods, Deep Learning methods have gained notoriety in recent years since they are showing better performance in tackling many problems of interest in the Machine Learning community (Krizhevsky et al., 2012; LeCun et al., 2015). The most classic and first successful application of such models was in optic character recognition (Lecun et al., 1998). In this previous study, a special case of an artificial neural network model known as a convolutional neural network (CNN) or fully convolutional network (FCN) was developed. The basic structure of such a model is the stack of convolutional and pooling layers followed by a series of fully connected layers and a final classification layer. The motivation for convolution and pooling operations is to include sparsity in the network, parameter sharing, and to account for small translational variances (Goodfellow et al., 2016). More recently, variations of FCN have been developed and applied in several areas from optical character recognition and image classification to medical imaging diagnostic (Krizhevsky et al., 2012; Redmon et al., 2016). However, FCN has the drawback of, generally, demanding extremely large datasets for the training process and to reduce the size of the image after each set of convolution and pooling operations. To address these drawbacks, adaptations and extensions for semantic segmentation have been developed. The models known as encoder-decoder networks have presented an improvement to the overall quality of the prediction by coupling the FCN to another set of

layers known as deconvolution network; examples of such models are U-Net and Segnet (Badrinarayanan et al., 2015; Ronneberger et al., 2015). The function of the deconvolution layers is to rescale the network output to the original image size without losing the information gained during the previous convolutional layers. However, SegNet has an advantage over other similar methods, in that the decoder layers share the same feature map indexes with the encoder layers. This design generates networks that have a high reduction in the number of parameters (Badrinarayanan et al., 2015).

Despite the fact that the Deep Learning models used are state of the art, they are still considerably complex with several layers in each encoder or decoder stack. Moreover, the difference in the size of the input image increased the network complexity because it directly affects the size of the input and output layers. Table 2.2 presents these differences in computational resources required between the networks evaluated for prediction of pixel classes in a single image. As the networks increase in complexity, it also will increase the demand for computational resources, more accentuated in RAM and also in processing time if using only a CPU. In the current application, the main factor driving the increased demand for computational resources was the size of the image used as input.

This analysis is important for the deployment of the CVS since it directly affects the hardware requirement. For instance, the model that has an input image 20% of the original size (390 x 520 pixels) and 4 encoder stacks had similar to better accuracy and IoU than more complex models (Figures 2.5 and 2.6). On the other hand, if the models were compared based on their computational performance, the selected model (input of 390 x 520 and 4 encoder stacks) uses on average 2 times less RAM and it is 4 times

faster on graphics processing unit (GPU) and uses 5 times less RAM and it is 5 times faster on the CPU than the more complex models. Therefore, this compact smaller model is more useful for deployment in embedded or mobile systems which can have a positive impact on the fish industry by potentially reaching more farmers. This concept had been already applied for other Deep Learning models (Lane and Georgiev, 2015) and there are also efforts from the scientific community to compress and improve the efficiency of such models (Kim et al., 2016). Thus, a more complex and robust model can possibly be deployed in embedded systems if it proved more accurate. Such a model may be important if there is a need to improve classification or to include different fish species in the model as well as different background scenarios. However, for the current application, the real interest is to correctly separate the Nile tilapia body for the estimation of body measurements and prediction of BW, CW, and CY, thus the lower IoU for fish fins is not of major concern. It is also worth noting that maybe the dataset used in this study is not large and diverse enough for training more complex models.

In the current study, we applied a classic data augmentation and batch normalization as a way to increase regularization, and improve learning performance and stability of the networks (Goodfellow et al., 2016). An approach to improve even more the training would be via other data augmentation techniques, such as synthetic data generated via rendering of computer graphics models (Abu Alhaija et al., 2018) or via generative adversarial networks (Frid-Adar et al., 2018), thus creating a more diverse, simulated dataset for training. Another technique for working with small datasets is transfer learning. The basic idea is to use a network model developed and trained in a larger and more complex data set for initialization of the weights, or as part of a new

model for a specific case. Examples in computer vision are FCN as ImageNet and VGG (Krizhevsky et al., 2012; Simonyan and Zisserman, 2015) that were developed using large and general purpose datasets as ImageNet, COCO, and PascalVoc (Deng et al., 2009; Everingham et al., 2015; Yang et al., 2015). These datasets are composed of hundreds of thousands of labeled images from diverse scenes and applications. The models developed for such general datasets have been fully adapted for other problems (Redmon et al., 2016; Zhou et al., 2017) or even used as part of the network in novel architectures. Therefore, the use of transfer learning and other data augmentation approaches may be useful for the case of including other fish species and different backgrounds.

2.5.2 Image Derived Body Measurements and Body Weight Prediction

It is well known that BW has a strong correlation with morphometric traits for many animal species. Previous work with Nile tilapia shows that phenotypic and genetic correlations between BW or CW with morphometric traits tend to be positive and high, while for CY with other traits tend to be positive but low to moderate (He et al., 2017; Turra et al., 2018). The results for phenotypic correlations presented in the current work (Table 2.3) agree with previous results for Nile tilapia, mainly with BA being the trait that achieved the best results. These correlation results are reflected in the predictive linear models evaluated, because the simple model that included only BA as a predictor variable achieved high predictions for BW and CW on the test dataset (Table 2.4). Therefore, the results showed the potential for the application of CVS in the Nile tilapia industry. The possibility of prediction of BW and CW on live animals without the need of weighing the

animals on a scale could speed-up the measurement process since several traits of interest can be measured at the same time with the proposed CVS. Regarding the predictions for CY, even though the MAE and the root mean square error (RMSE) were low, the R^2 was also low. These results suggest that while it is possible to achieve high R^2 for BW and CW with linear models, there is still room for improvement regarding the prediction of their ratio, CY. In addition, it is possible that CY does not have a linear relationship with the variables evaluated, and additional variables should be tested for improvement of CY prediction.

It is important to note that the developed methodology is applicable only to Nile tilapia, and the images need to be taken using a homogeneous green background. As an ultimate goal, it would be interesting to have a CVS where there is no need for handling the animals, with the images been acquired underwater. Such an idea is not new and there are some attempts that have been made for other fish species, with the more successful ones for Tuna (Muñoz-Benavent et al., 2018; Shafait et al., 2017). However, all those applications use restrictive body shape models and thus can only be applied for mature tuna. Meanwhile, we predict that the image segmentation methodology described in this study can be extrapolated to underwater images and also for other species. The main constraint in adapting the image segmentation described in this study for other fish species is the lack of a database rich enough for training such a model.

2.6 Conclusion

In the current paper, we presented the application of Deep Learning models for the development of a computer vision system for the measurement of live Nile tilapia

morphometric traits and prediction of BW and CW. The current method achieved high predictive accuracy, with R^2 of 0.95 and 0.94 for BW and CW, respectively, while measuring other traits of interest at a satisfactory speed (0.21 seconds for evaluation of a single image for the best segmentation model). However, the current CVS is still not general enough to be applied to other fish species or with more texturized backgrounds. In addition, CY was poorly predicted, showing a very low R^2. Hence, other strategies need to be devised to achieve better predictive ability for this specific trait.

2.7 References

Abu Alhaija, H., Mustikovela, S.K., Mescheder, L., Geiger, A., Rother, C., 2018. Augmented Reality Meets Computer Vision: Efficient Data Generation for Urban Driving Scenes. Int. J. Comput. Vis. 126, 961–972. https://doi.org/10.1007/s11263-018-1070-x

Badrinarayanan, V., Kendall, A., Cipolla, R., 2015. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. arXiv.

Calcagno, V., Mazancourt, C. de, 2010. glmulti : An R Package for Easy Automated Model Selection with (Generalized) Linear Models. J. Stat. Softw. 34, 1–29. https://doi.org/10.18637/jss.v034.i12

CONCEA, 2016. Normativas do CONCEA para produção, manutenção ou utilização de animais em atividades de ensino ou pesquisa científica, 3rd ed. Conselho Nacional de Controle de Experimentação Animal, Brasília, Brazil.

Costa, C., Antonucci, F., Boglione, C., Menesatti, P., Vandeputte, M., Chatain, B., 2013. Automated sorting for size, sex and skeletal anomalies of cultured seabass using external shape analysis. Aquac. Eng. 52, 58–64. https://doi.org/10.1016/J.AQUAENG.2012.09.001

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. ImageNet: A Large-Scale

Hierarchical Image Database, in: IEEE Computer Vision and Pattern Recognition (CVPR). p. 8.

Everingham, M., Ali Eslami, S.M., Van Gool, L., K I Williams, C., Winn, J., Zisserman, A., 2015. The PASCAL Visual Object Classes Challenge: A Retrospective. Int J Comput Vis 111, 98–136. https://doi.org/10.1007/s11263-014-0733-5

FAO, 2018. The State of World Fisheries and Aquaculture 2018 - Meeting the sustainable development goals. FAO, Rome.

Fernandes, A.F.A., Silva, M. de A., Alvarenga, E.R. de, Teixeira, E. de A., Silva Junior, A.F. da, Alves, G.F. de O., Salles, S.C.M. de, Manduca, L.G., Turra, E.M., 2015. Morphometric traits as selection criteria for carcass yield and body weight in Nile tilapia (*Oreochromis niloticus L.*) at five ages. Aquaculture 446, 303–309. https://doi.org/10.1016/j.aquaculture.2015.05.009

Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., Greenspan, H., 2018. Synthetic data augmentation using GAN for improved liver lesion classification, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, pp. 289–293. https://doi.org/10.1109/ISBI.2018.8363576

Gjedrem, T., Robinson, N., Rye, M., 2012. The importance of selective breeding in aquaculture to meet future demands for animal protein: A review. Aquaculture 350–353, 117–129. https://doi.org/10.1016/J.AQUACULTURE.2012.04.008

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press, Cambridge, MS.

He, J., Zhao, Y., Zhao, J., Gao, J., Han, D., Xu, P., Yang, R., 2017. Multivariate random regression analysis for body weight and main morphological traits in genetically improved farmed tilapia (*Oreochromis niloticus*). Genet. Sel. Evol. 49, 80. https://doi.org/10.1186/s12711-017-0357-7

Jeong, S.-J., Yang, Y.-S., Lee, K., Kang, J.-G., Lee, D.-G., 2013. Vision-based Automatic System for Non-contact Measurement of Morphometric Characteristics of Flatfish. J. Electr. Eng.

Technol. 8, 1194–1201. https://doi.org/10.5370/JEET.2013.8.5.1194

Kim, Y.-D., Park, E., Yoo, S., Choi, T., Yang, L., Shin, D., 2016. Compression of Deep Convolutional Neural Networks for fast and low power mobile applications, in: International Conference on Learning Representations. p. 16.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet Classification with Deep Convolutional Neural Networks, in: Neural Information Processing Systems. pp. 1097–1105.

Lane, N.D., Georgiev, P., 2015. Can Deep Learning Revolutionize Mobile Sensing?, in: Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications - HotMobile '15. ACM Press, New York, New York, USA, pp. 117–122. https://doi.org/10.1145/2699343.2699349

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539

Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86, 2278–2324. https://doi.org/10.1109/5.726791

Muñoz-Benavent, P., Andreu-García, G., Valiente-González, J.M., Atienza-Vanacloig, V., Puig-Pons, V., Espinosa, V., 2018. Enhanced fish bending model for automatic tuna sizing using computer vision. Comput. Electron. Agric. 150, 52–61. https://doi.org/10.1016/J.COMPAG.2018.04.005

Navarro, A., Lee-Montero, I., Santana, D., Henríquez, P., Ferrer, M.A., Morales, A., Soula, M., Badilla, R., Negrín-Báez, D., Zamorano, M.J., Afonso, J.M., 2016. IMAFISH_ML: A fully-automated image analysis software for assessing fish morphometric traits on gilthead seabream (Sparus aurata L.), meagre (Argyrosomus regius) and red porgy (Pagrus pagrus). Comput. Electron. Agric. 121, 66–73. https://doi.org/10.1016/j.compag.2015.11.015

R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You Only Look Once: Unified, Real-Time

Object Detection.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Navab, N., Hornegger, J., Wells, W., Frangi, A. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Springer, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Shafait, F., Harvey, E.S., Shortis, M.R., Mian, A., Ravanbakhsh, M., Seager, J.W., Culverhouse, P.F., Cline, D.E., Edgington, D.R., 2017. Towards automating underwater measurement of fish length: a comparison of semi-automatic and manual stereo–video measurements. ICES J. Mar. Sci. 74, 1690–1701. https://doi.org/10.1093/icesjms/fsx007

Simonyan, K., Zisserman, A., 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv.

The MathWorks, 2017. MATLAB Release 2017b.

Turra, E.M., Fernandes, A.F.A., de Alvarenga, E.R., Teixeira, E.A., Alves, G.F.O., Manduca, L.G., Murphy, T.W., Silva, M.A., 2018. Longitudinal analyses of correlated response efficiencies of fillet traits in Nile tilapia. Animal 12, 445–453. https://doi.org/10.1017/S1751731117001768

Wickham, H., 2009. ggplot2 Elegant Graphics for Data Analysis. Springer, New York, NY. https://doi.org/10.1007/978-0-387-98141-3

Yang, L., Zhang, L., Dong, H., Alelaiwi, A., Saddik, A. El, 2015. Evaluating and Improving the Depth Accuracy of Kinect for Windows v2. IEEE Sens. J. 15, 4275–4285. https://doi.org/10.1109/JSEN.2015.2416651

Zhou, S.K., Greenspan, H., Shen, D., 2017. Deep learning for medical image analysis, Deep Learning for Medical Image Analysis. Elsevier.

Chapter 3: A novel automated system to acquire biometric and morphological measurements and predict body weight of pigs via 3D computer vision

3.1 Abstract

Computer vision applications in livestock are appealing since they enable measurement of traits of interest without the need to directly interact with the animals. This allows the possibility of multiple measurements of traits of interest with minimal animal stress. In the current study, an automated computer vision system was devised and evaluated for extraction of features of interest, such as body measurements and shape descriptors, and prediction of body weight in pigs. From the 655 pigs that had data collected 580, had more than 5 frames recorded and were used for the development of the predictive models. The cross-validation for the models developed with data from nursery and finishing pigs presented an R^2 ranging from 0.86 (random selected image) to 0.94 (median of images truncated on the third quartile), whereas with the dataset without nursery pigs, the R^2 estimates ranged from 0.70 (random selected image) to 0.84 (median of images truncated on the third quartile). However, overall the mean absolute error was lower for the models fitted without data on nursery animals. From the body measures extracted from the image, body volume, area, and length were the most informative for prediction of body weight. The inclusion of the remaining body measurements (width and heights) or shape descriptors to the model promoted significant

improvement of the predictions, whereas the further inclusion of sex and line effects were not significant.

3.2 Introduction

It is of great importance in pig production to define efficient management strategies in order to achieve optimum pig marketing weight and grade size in the most cost-efficient fashion. However, optimization of feeding and management in growing-finishing pig units is difficult (de Lange et al., 2001). Accurate prediction of the growth curve of a pig and understanding the factors affecting growth are valuable tools to assist producers in improving the efficiency of pork production (de Lange et al., 2001; Gous et al., 2006). Frequent measures of pig body weight (BW) can enable precise estimation of an individual animal's growth curve and assessment of inter and intra-group variability. However, consecutive manual measurements of BW are not feasible in practice since it is labor-intensive and costly, and may increase animal stress, leading to reduced animal performance and even animal loss (Faucitano and Goumon, 2018; Grandin and Shivley, 2015). In order to acquire frequent measures of BW, one alternative is via the use of automatic scales installed in each pen. However, these scales represent a physical intervention in the environment and the animals need an adaptation period (Kongsro, 2014). Moreover, such scales can be expensive. On the other hand, computer vision systems (CVS) offer a non-invasive approach, via the use of strategically positioned cameras. Thus, the use of CVS could be a powerful technology to assess real-time growth curves with minimal negative impact. Recently, CVSs based on depth cameras have been proposed to predict pig BW (Condotta et al., 2018; Kongsro, 2014). However, these

applications involve some level of manual processing either for selection of the best images out of the group of captured frames, image segmentation, features extraction, among others. Since manual processing is not feasible for a large-scale application, full automation is a key factor in order to deploy an optimized CVS at commercial settings. Therefore, the objectives of this study were: 1) to develop an automated CVS for depth video processing, and extraction of phenotypes of interest, and 2) to evaluate the ability of the models developed to predict pig body weight at commercial farm conditions.

3.3 Material and Methods

The data set of animal BW and recordings of the weighing process were supplied by the Pig Improvement Company (PIC, a Genus company, Hendersonville, TN). PIC follows rigorous animal handling procedures that are in compliance with federal and institutional regulations regarding proper animal care practices (FASS, 2010).

3.3.1 Local, Animals and Devices

The study was conducted in a pig multiplier facility in the United States. A total of 655 animals were randomly selected including males and females from 3 different terminal lines. From those pigs, 37 were nursery animals with an average body weight BW of 32 kg, and 618 were late finishing pigs with an average BW of 120 kg. The video acquisition was performed using a Kinect V2 sensor (Microsoft, Redmond, WA) which has an RGB camera (resolution of 1920 × 1080 pixels), depth sensors (resolution of 512 × 424 pixels), and microphone array. The video emulation, processing, feature extraction,

and training of predictive models were performed with a Windows 10 computer equipped with an Intel i7-7700k CPU (Intel, Santa Clara, CA) and an NVIDIA Quadro M4000 GPU (NVIDIA, Santa Clara, CA).

3.3.2 Data Acquisition

In this study, groups of pigs that were housed in the same pen were moved to a separate area where they were individually weighed. The difference from a traditional weighing process, was the introduction of the CVS for image acquisition in the area preceding the scale and the fact that all pigs from a pen were weighed. The electronic scale used was an EziWeigh5i (Tru-Test, Mineral Wells, TX), which has a measured standard error of ± 1% of the load. Before weighing the pigs, the scale was calibrated using blocks of known weight. The period each pig stayed in the prescale area was random and dictated by the time required to measure and annotate the weight of the previous pig in the line.

The recording process was initialized before the first pig of each group was weighed. Thus, each group had its own video with respective animal order and weight. The RFID sensor (radio-frequency identification) from each animal was used in order to obtain the pig identification (ID) and order in which they entered the prescale area. The pig ID was also manually annotated. This information was used to evaluate the CVS ability to recognize animals and keep track of the order information.

Subsequently, the videos were automatically processed for image segmentation and features extraction for each animal. To reproduce an actual implementation of the CVS, the video and the segmentation algorithm were processed in two parallel

independent routines. The first was the video emulation of an actual Kinect camera via the Kinect for Windows SDK v2.0 (Microsoft, 2014). The second was the CVS containing the segmentation algorithm that retrieves and processes frames from a connected Kinect camera. The CVS presented in this study corresponds to a library of custom codes written in MATLAB (Release 2017b) (The MathWorks, 2017). The connection between the CVS and the emulated video was made available by internal functions from the Kinect for Windows SDK (Microsoft, 2014). For the efficient establishment of this connection, the Kinect for Windows SDK custom C++ code was encapsulated within a MATLAB MEX function format following the directions in the Kin2 toolbox (Terven and Córdova-Esparza, 2016). Therefore, it was possible to connect the emulated Kinect device and retrieve the sensor intrinsic parameters, as the depth sensor effective focal length (f). To minimize the noise on the set of segmented images, the segmentation algorithm also classified each frame as acceptable or not based on animal posture (Figure 3.1). If the frame was accepted, then its features were extracted and saved for evaluation. The segmentation process was split into 3 steps: 1) first segmentation (recognize if there is a pig in the image); 2) second segmentation (segment the body of the animal in the image), and 3) feature extraction (extract the features of interest related to BW). Details on each of these steps are provided below.

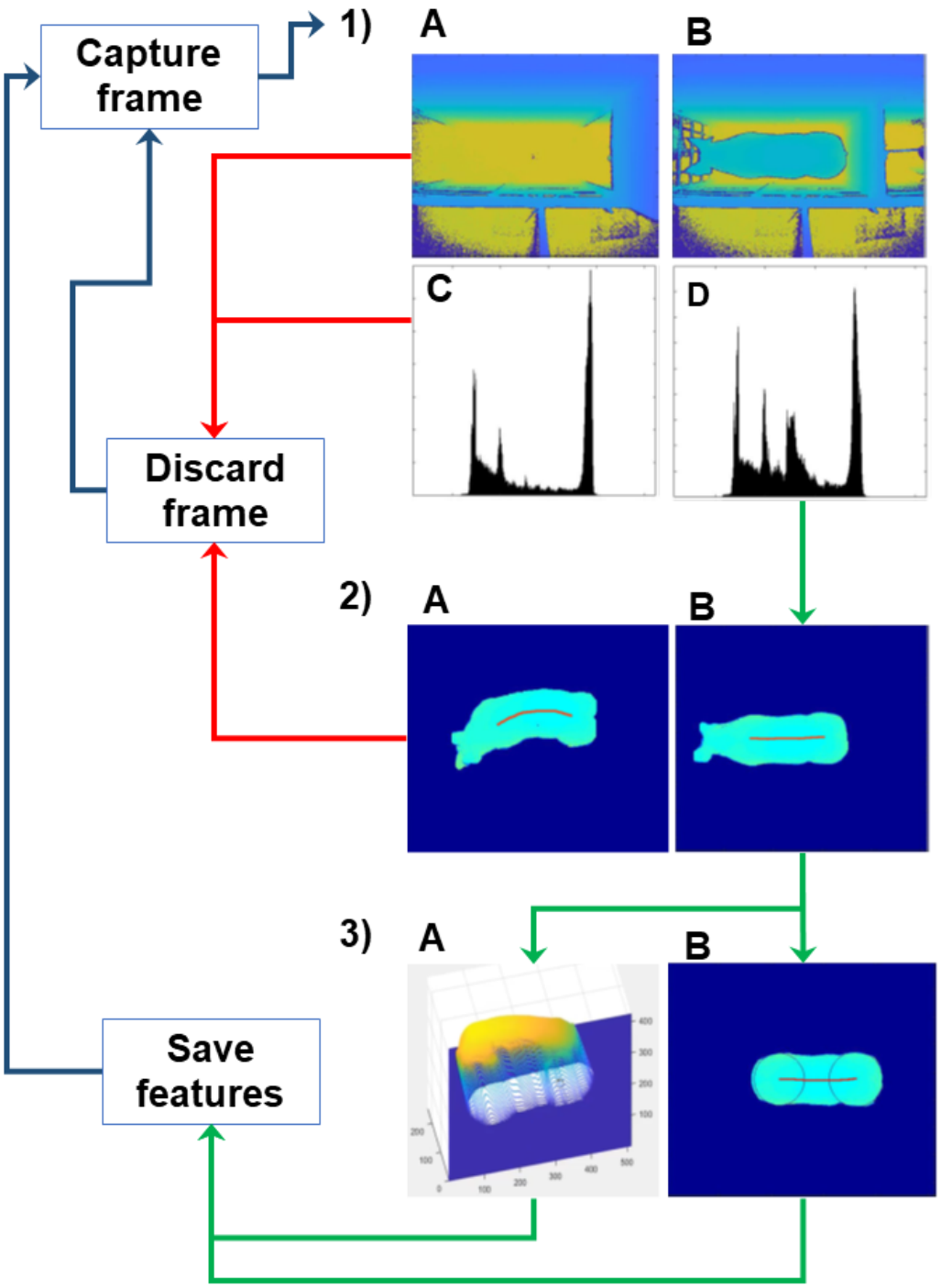

Figure 3.1 *Diagram of the computer vision algorithm devised:* ***1)*** *First segmentation,* ***A*** *is a frame without the pig,* ***B*** *a frame with a pig, and* ***C*** *and* ***D*** *are their respective histograms.* ***A*** *is used to acquire the distance to the floor and then discarded, whereas* ***B*** *is passed to the next step.* ***2)*** *Second segmentation:* ***A*** *discards current frame if the pig is touching the border and/or if it is not with a straight posture,* ***B*** *otherwise removes the pig head and tail from the image and passes the image to the next step.* ***3)*** *Rotates and centralizes the image, and then performs estimation of body traits and shape descriptors. The estimated features for the current pig frame are finally saved. After the current frame is fully processed or discarded, the system captures the next available frame from the connected Kinect device.*

3.3.3 First Segmentation

This step corresponds to the identification of the scene presented on the image. Basically, it differentiates whether the scene presented in the image contains a pig. To identify if there was a pig on a frame, the depth matrix was processed as an intensity image. In this format, the intensity value in each pixel corresponds to the actual distance from the camera. Then, each pixel in the intensity image can be classified into background or foreground (objects in the scene). For this classification process, an adaptive threshold with a sensitivity of 0.4 was applied (Bradley and Roth, 2007). The largest foreground object identified was considered as the pig (Figure 3.2-C). Before the first pig was present, the algorithm performed an estimation of the distance between the camera and floor based on the mode of the group of pixels farthest from the camera (equation 3.1):

$$fdist = mode(maxPeak), \quad (1)$$

where $fdist$ is the estimated distance from the floor to the camera and $maxPeak$ is the cluster of pixels above the 90^{th} percentile of pixel distance (Figure 3.1-C). After identification of the pig on the image, the CVS passes the frame to the second segmentation. Otherwise, after estimation of the distance from the camera to the floor, the program starts over with the next available frame.

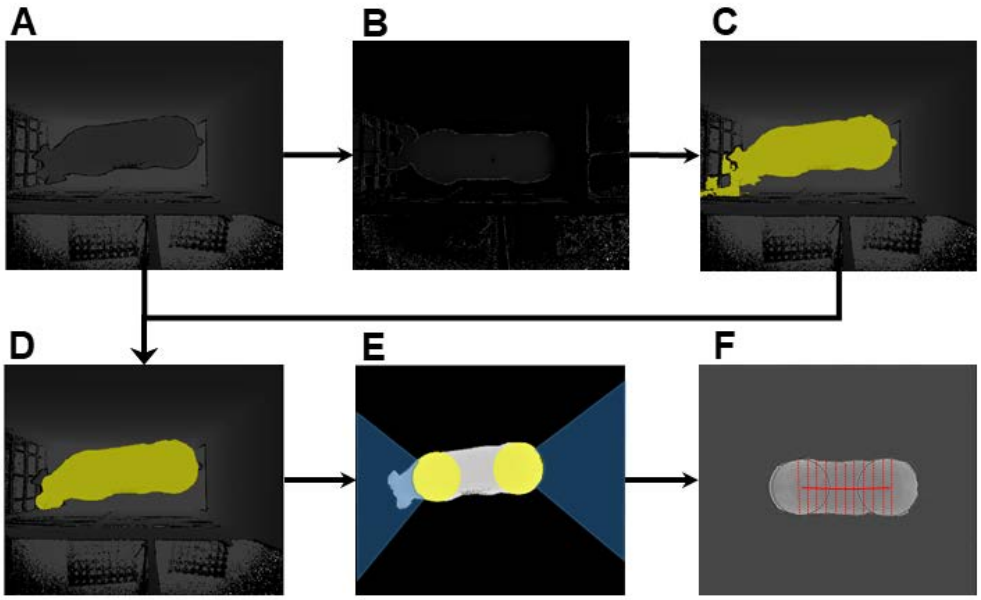

Figure 3.2 *Image processing within the first and second segmentation.* ***A:*** *Depth frame from a Kinect sensor.* ***B:*** *Resulting image after adaptive threshold applied.* ***C:*** *Mask of the object selected as the pig.* ***D:*** *Cleaning and smoothing of the mask via image opening.* ***E:*** *Identification of shoulders and rump via Hough transform and removal of remaining head and tail as the regions adjacent to the shoulder and rump.* ***F:*** *Rotated and centralized pig with the delimited predicted spine (vertical line) and widths (horizontal lines).*

3.3.4 Second Segmentation

After a pig was identified on the scene, a series of criteria were checked in order to accept the frame as valid. This checking is important in order to remove frames in which pigs were either not well positioned or the image processing algorithm did not perform as well as intended. The first criterion to select the frame was to identify if the pig was not connected to the border of the scene. This criterion was evaluated via the morphological opening of the logical image derived from the adaptive threshold (Figure 3.2-D). In this step, every object connected to the border of the region of interest was removed from the logical image. For example, if a pig was pressed against a wall or door, the correspondent frame was discarded. Also, the pig body position was evaluated by an outline of the pig dorsal area and estimation of its spine position. For estimation of the spine position, only

the back area that was between the shoulder and rump of the pig was used. Shoulder and rump were identified via an adaptation of Hough transform to identify round objects (Atherton and Kerbyson, 1999; Hough, 1962). The Hough transform is a method initially developed for identification of lines in the picture plane of a grayscale image. The method consists of the identification of edge points and the assignment of an indicator value to each accumulator corresponding to a specific line position and orientation. The highest peaks in the accumulator array correspond to the strongest lines (Szelisk, 2011). The pig head and tail were identified and removed as the regions in front of the shoulder and after the rump (Figure 3.2-E). Spine “backbones” were estimated as the median points across sections of the thoracic and lumbar area estimated from the dorsal view. A total of 11 points, approximately every other backbone on the thoracic (14 to 15 vertebras) and lumbar (6 to 7 vertebras) regions were estimated. To ensure a more robust prediction of the spine curvature a cubic curve was fitted over those points and used as the predicted back spine. A pig was considered as having a straight position if the spine curvature approximated a straight line by a linear equation with a coefficient of determination R^2 of 0.95 or higher. If a frame achieved these criteria, it was passed to the feature extraction step. Otherwise, the current frame was discarded and the program started over with the next available frame.

3.3.2 Feature Extraction

Finally, for any accepted frame, the algorithm rotates and centers the pig in the image and then performed the feature extraction and saved the original frame, as a black and white mask of the pig and the corresponding extracted features of interest. The set

of extracted features was divided into 2 classes: body measurements and shape descriptors.

The body measurements were: area, volume, length, 11 widths (W1 to W11), and 11 heights (H1 to H11) at equidistant locations across the back of the animal, from the shoulders to the rump. The length, width, and area of an object on an image can be estimated by the sum of its pixels. However, the estimation of the real measurement is also affected by the distance from the camera to the object (i.e., the same object will have a smaller area in the formed image if positioned farther from the camera). To estimate the real area of any pixel, the trigonometric principle of image magnification can be used:

$$m = f/d_o \tag{2}$$

$$a_p = 1/m \tag{3}$$

$$a_o = \sum_{p=1}^{n} a_p \tag{4}$$

$$v_o = \sum_{p=1}^{n} a_p * h_p \tag{5}$$

where m is the magnification factor, f is the effective focal length that is intrinsic to the camera, and d_o is the distance the object is from the camera (equation 3.2). The area of a pixel a_p in the metric scale, on the object plane, can be approximated in this case from the magnification factor as shown in equation 3.3. Thus, the area of any object a_o can be estimated as the sum of the area of the pixels a_p that form its image (equation 3.4). The index *p* (p = 1, …, n) refers to the n pixels that form the image representation of the pig. The pixel heights h_p can be estimated by the difference between the distance to the camera for the pixel p in the pig dorsal area and the distance from the camera to the floor.

Therefore, the animal volume v_o is the sum of the volumes specific to each pixel that constitutes the image representation of the pig (equation 3.5). It is important to point out that this volume is not the real volume of the animal body, but an apparent volume since only the top view of the animal was obtained, and then projected to the floor.

The shape descriptors used in the present study were 1) eccentricity, which is a measurement of roundness, and was estimated as the ratio between the foci and the major axis of the ellipsis that has the same second moments as the pig area, and 2**)** the first 20 Fourier descriptors. Fourier descriptor is a class of global image descriptors normally used for shape analysis and image matching (Bowman et al., 2000; Burger and Burge, 2016; Zhang and Lu, 2002). For robustness, each segmented pig image was rotated, centralized and transformed to the polar plane representation previous to the estimation of its discrete Fourier transform in 1 dimension. Thus, the Fourier descriptors for the polar image can be estimated via a discrete Fourier transformation as follows:

$$FD_k = \sum_{j=0}^{l-1} X_j w_l^{jk} \quad , \tag{6}$$

where FD_k $(k = 0, 1, \dots)$ is each of the Fourier descriptors for the signal *X* of length $\boldsymbol{l}$ (the number of columns of the polar image) and $w_l = e^{-(2\pi i)/l}$, with $i = \sqrt{-1}$ the imaginary constant. Since FD_0 (the first Fourier descriptor) corresponds to the sum over the signal, i.e. the area of the object in the image, all Fourier descriptors were also divided by FD_0 for invariance on animal size.

After processing a given frame, the algorithm moves to the next available frame. To count the number of pigs and assign each video frame to the correct animal, the

following criterion was used: If there was a sequence of frames without a pig or if there was a pig appearing in the opposite location from the previous recorded, the next valid frame was assigned to the next animal. Thus, for each video session, the number of pigs and the number of frames for each pig were tracked, frames with their features were automatically assigned to the respective pig, and the dataset generated from each video session was saved for subsequent statistical analysis.

3.3.3 Statistical Analysis

The dataset of extracted image features and measured body weights was merged and manipulated for statistical analysis in R (R Core Team, 2016) via custom code from the authors. Since the number of frames generated per animal during the image acquisition can be large (up to 500 frames), strategies to reduce data dimension were evaluated. Each reduced dataset had just 1 data point for each variable for each animal. Before frame selection to build the reduced dataset, outlier frames from each individual animal were removed. Eight strategies to reduce data dimension were evaluated as follows: 1) random selection of a single image from each pig (Random); 2) selection of the features from the image with highest estimated animal area (Max Area); 3) volume (Max Volume); 4) length (Max Length); 5) computing the average for every feature across all frames for each pig (Average); 6) computing the truncated average with exclusion on 20% of the extreme data for each animal (Truncated Average); 7) computing the median for every feature across all frames for each pig (Median); and 8) computing the truncated median for every feature across of the subset at the third quartile of the extracted variables (third Quartile). Notice that the truncated average and third quartile methods are

calculating averages across a reduced number of images. For an animal with few images, for example, 5 original images, this subset will have 3 images only, after the average or median is calculated. These reduced datasets were used to predict the BW through linear models using the following equation:

$$\boldsymbol{y} = \sum_{j=1}^{k} \beta_j \boldsymbol{x}_j + \boldsymbol{e} \tag{7}$$

where $\boldsymbol{y}$ is the vector of observations, β_j is the coefficient for the j predictor variable, $\boldsymbol{x_j}$ the vector of observations for the j predictor variable, and $\boldsymbol{e}$ is the vector of residual effects. For all the reduced datasets, 10 permutations on a 5-fold cross-validation (CV) were used to access the stability of the estimates. Also, within each step of the CV, a stepwise regression approach with the Akaike Information Criterion (AIC) used as model selection criterion was applied to access the importance of the predictor variables. As such, 400 different model selection scenarios were considered (8 reduced datasets × 10 permutations × 5 CVs). Stepwise regression was implemented using the *stepAIC* function of the MASS package in R (R Core Team, 2016; Venables and Ripley, 2002).

After evaluating the methods for data reduction and frequency of appearance of variables, 5 final models to predict pig BW were evaluated on the best data reduction scenario. These final models differ by the systematic inclusion of selected variables. The models included as follows: 1) Volume, Area, and Length (VAL); 2) VAL, Eccentricity and selected Polar Fourier descriptors (VALS); 3) VAL and selected heights and widths (VALB); 4) A model including all variables in VALS and VALB (VALBS); and 5) VALBS

plus sex and line effects (VALBSSL). To evaluate the prediction quality of these models, a 5-fold CV over 10 permutations was performed.

3.4 Results

The Kinect sensor records depth data at a rate of 30 frames per second (fps), whereas the devised CVS performed at a rate of 8 fps. Therefore, if a pig stayed under the sensor for 30 s (which was the average time for weighing an animal), it had a total of 240 frames evaluated by the CVS (30 s × 8 fps). However, the number of recorded frames for each animal was heterogeneous (Figure 3.3) because the recording time for each pig was varied. It is important to point out that the data were collected in a commercial setting, without much control and, because of this, some animals had just a few frames collected.

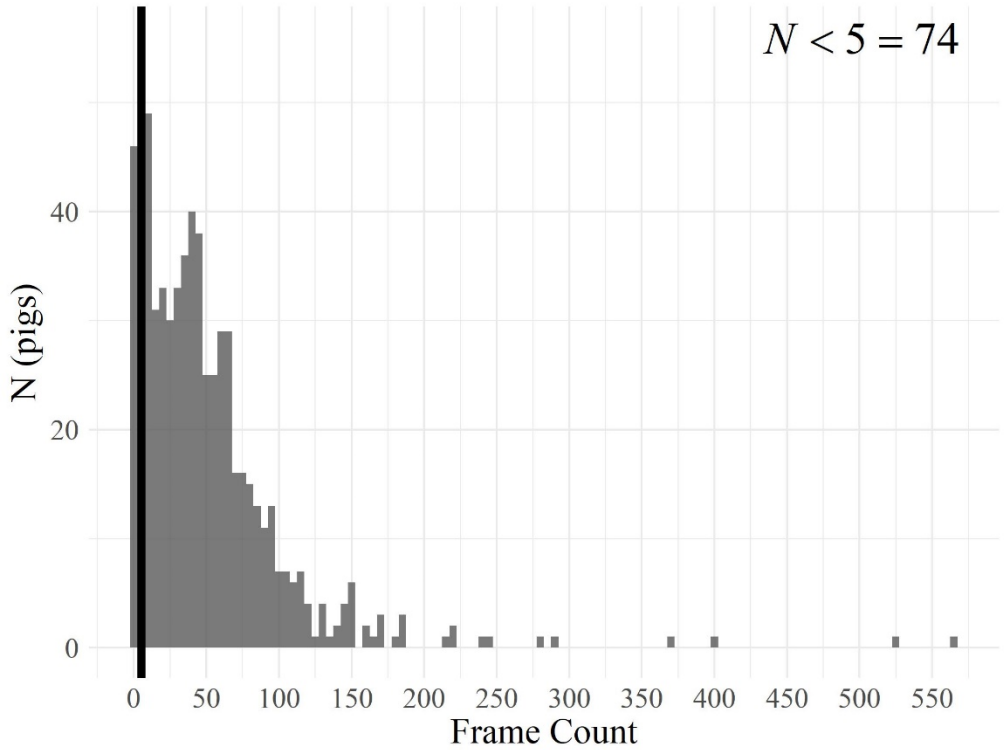

Figure 3.3 *Histogram of the number of frames (N) selected per pig. The black vertical line marks the number of pigs that had a count of 5 frames.*

The most common reasons for missing frames were as follows: 1) animals passing too quickly over the area; 2) animals that did not stay at an appropriate position; and 3) animals that passed together under the camera resulting in scene occlusion and consequently fewer to no valid frames for those animals. For the purpose of model training, animals with less than 5 frames recorded were removed from the dataset. Thus, a total of 580 pigs had their data analyzed in the present study. From this total, 21 were nursery animals with average BW of 31.5 kg and 559 finishing pigs with average BW of 119.8 kg (Figure 3.4).

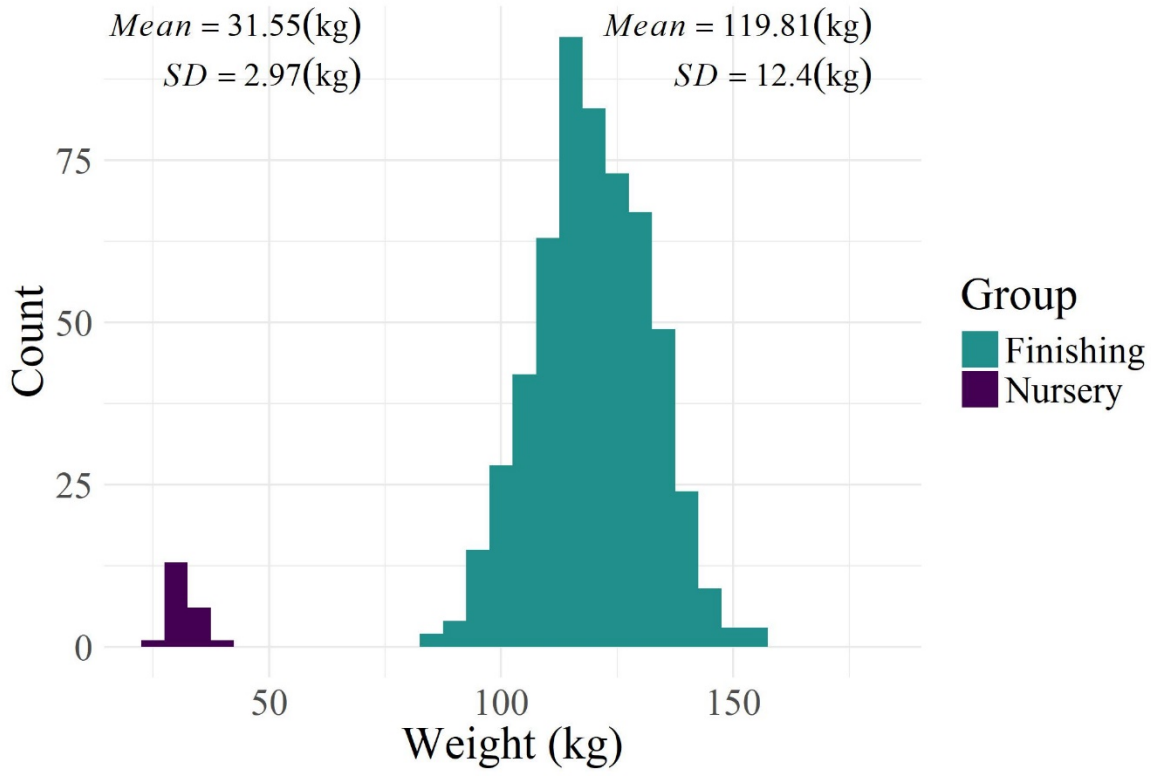

Figure 3.4 *Histogram of live body weight (kg) distribution for nursery and finishing pigs, with respective averages and standard deviations (SD).*

3.4.1 Analysis Including Nursery Data

A 5-fold CV was performed over 10 permutations for each reduced dataset; thus, a total of 400 model selections were performed (10 permutations on 5-fold CV and 8 different reduced datasets). The mean absolute error (MAE) and the coefficient of determination (R^2) for the selected models were estimated in the CV. A total of 265 different models were selected across the 400 different model selection scenarios (Figure

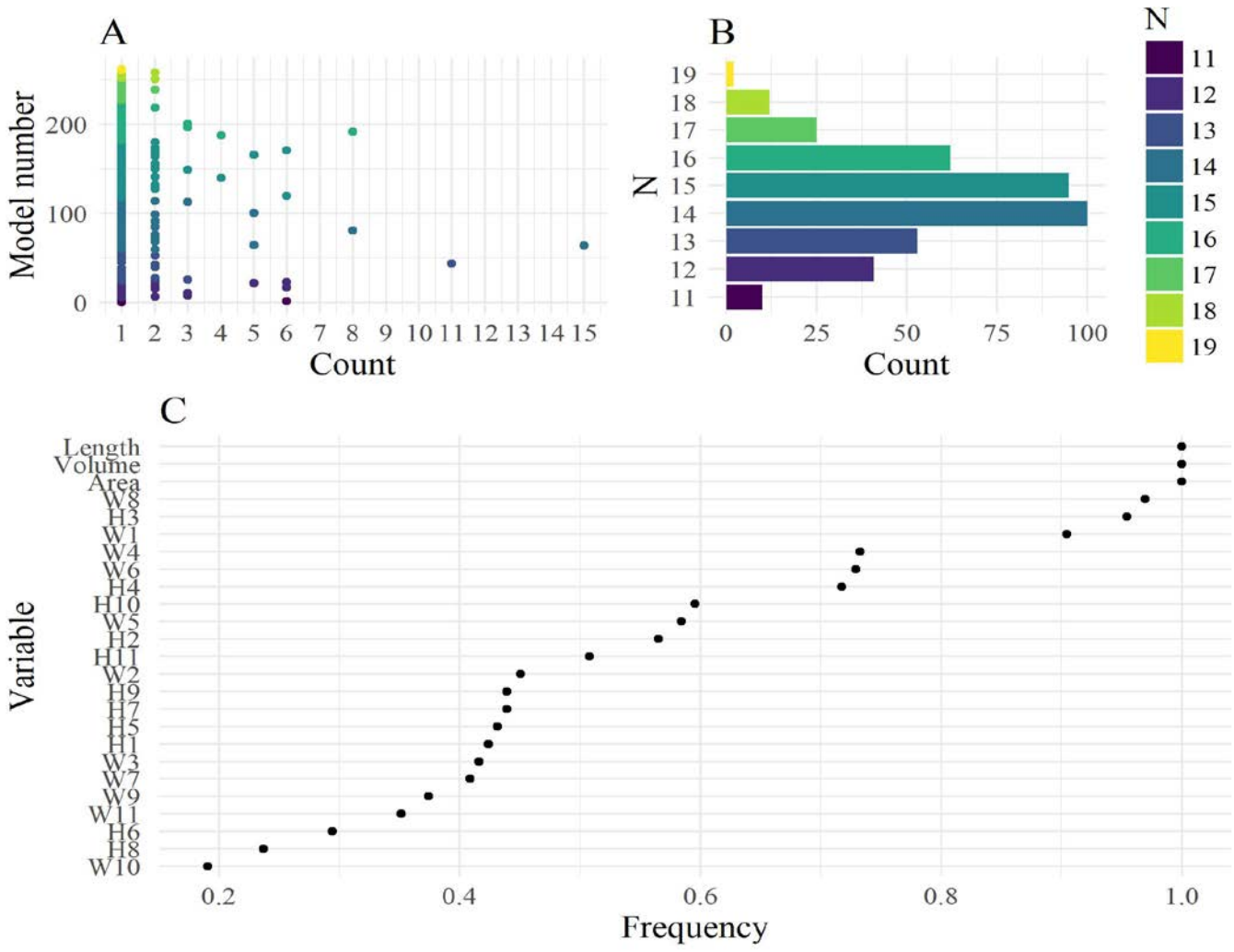

Figure 3.5 *Evaluation of stepwise model selection over the cross-validations for the 8 different reduced dataset methods when including the data on nursery and finishing pigs, and body measures importance where W1 to W11 and H1 to H11 refer to the 11 widths and heights measured. (**A**) Number of times (Count) a specific model (group of body measurement) was selected. (**B**) Histogram of the number of different models (Count) per number of variables (body measurements) selected in the model (N). (**C**) The frequency that each body measurement was selected across all selected models.*

3.5). Also, most of the selected models had 14 variables, whereas the single model that appeared the most, appeared 15 times and had 13 variables. Since there was a high correlation between body measures that are close to each other (e.g., H6 and H7), the several different models presented subtle switches between similar variables (Figure 3.6). It is worth noting that the model that contained all possible 25 variables was never selected, and the most complex models had 19 variables. From the set of variables selected in each validation round for each of those reduced datasets we can observe that body area, volume, and length were present in every model (Figure 3.5). Also, apart from those variables, the 3 models that appeared the most also presented in common 2 widths (W1, W8), and 2 heights (H3 and H7).

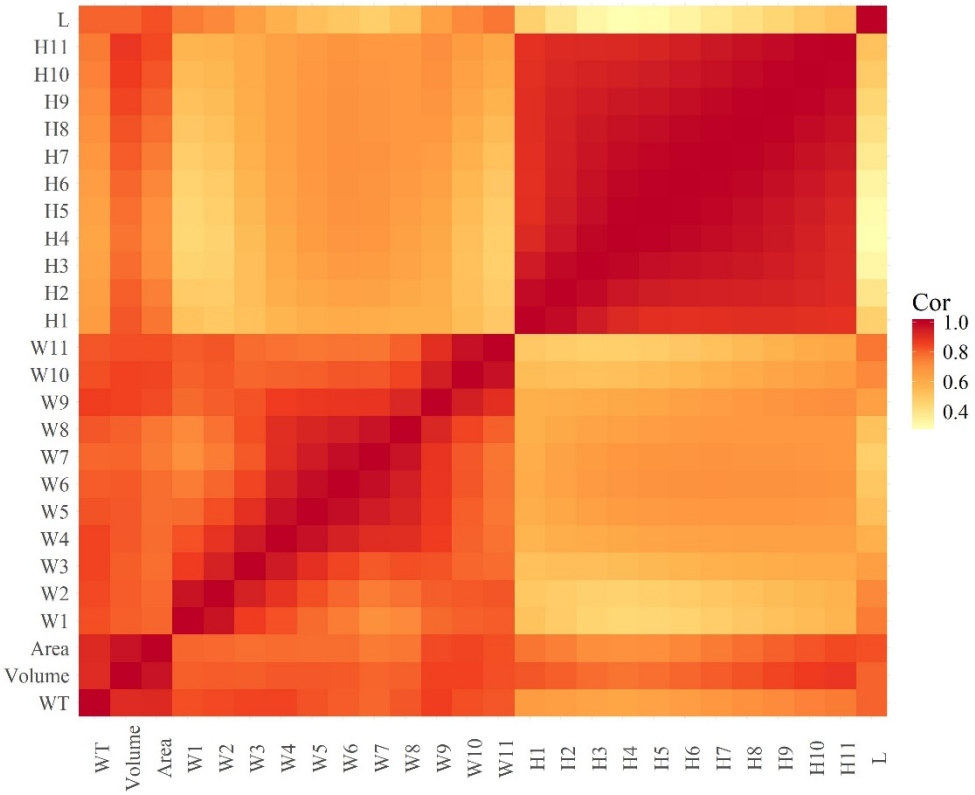

Figure 3.6 *Correlations (Cor) between body weight (BW), volume, area, widths (W1 to W11), heights (H1 to H11), and length for nursery and finishing animals.*

For each step of the CV and permutation, the MAE and R^2 for the selected model were estimated (Figure 3.7). This allows for a robust comparison between the possible ways to reduce the dataset. For all different reduced datasets, the R^2 was high with averages ranging from 0.87 for a randomly selected image for each animal to 0.925 for the average of the image variables at the third quartile for each animal. However, there is a separation between using the information from a single image against summarizing the information from several images of the same animal. Thus, the results for the reduced datasets that are originated from averages or median of the full sets of images for each animal performed better than the random selection of a single image, or the selection of the image that presented maximum area, length, or volume estimate for the given animal.

3.4.2 Analysis Without Nursery Animals

In this section, the nursery animals were removed from the dataset with the intent to evaluate if it is possible to improve the BW prediction of the heavier animals alone.

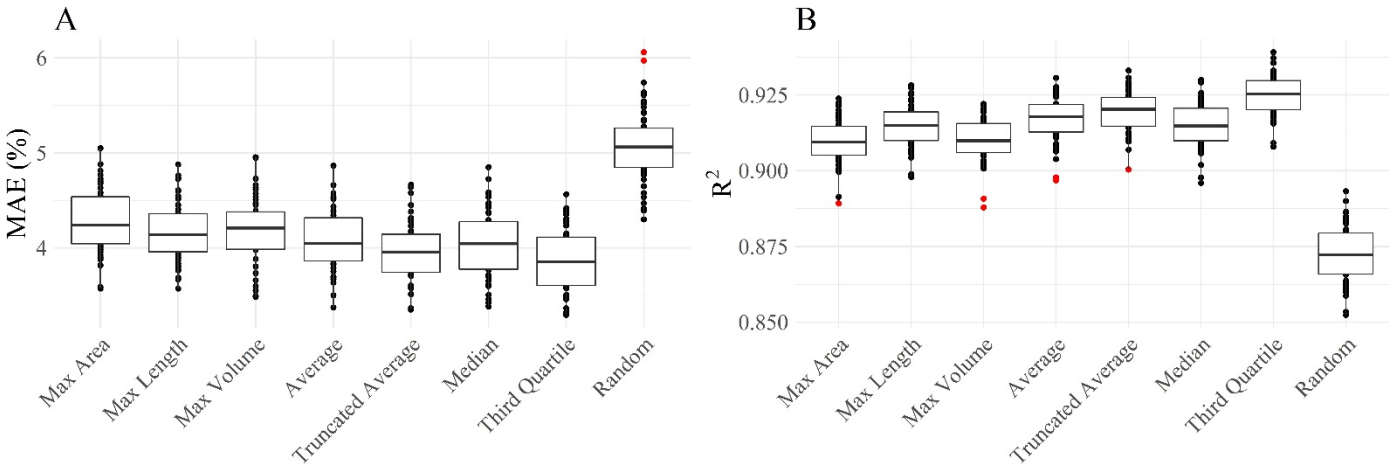

Figure 3.7 *Results for the different models across cross-validation for the dataset with nursery and finishing pigs. (**A**) Box plots for mean absolute error (MAE), as a percentage of the average body weight. (**B**) The coefficient of determination (R^2) of the different models across cross-validation sets. The hinges represent the 1.5 interval interquartile (approximately 95% confidence interval), and red dots outlier results.*

Initially, a similar analysis to the one in the previous section was conducted. That is, to

access the difference in data manipulation previously to fitting a linear model for prediction of BW (Figure 3.8-A and B). The results were similar to those from the data including nursery animals. The main difference was the reduction in the magnitude of the R^2. Previously, the calculated R^2 values ranged from 0.88 for the data set composed of a

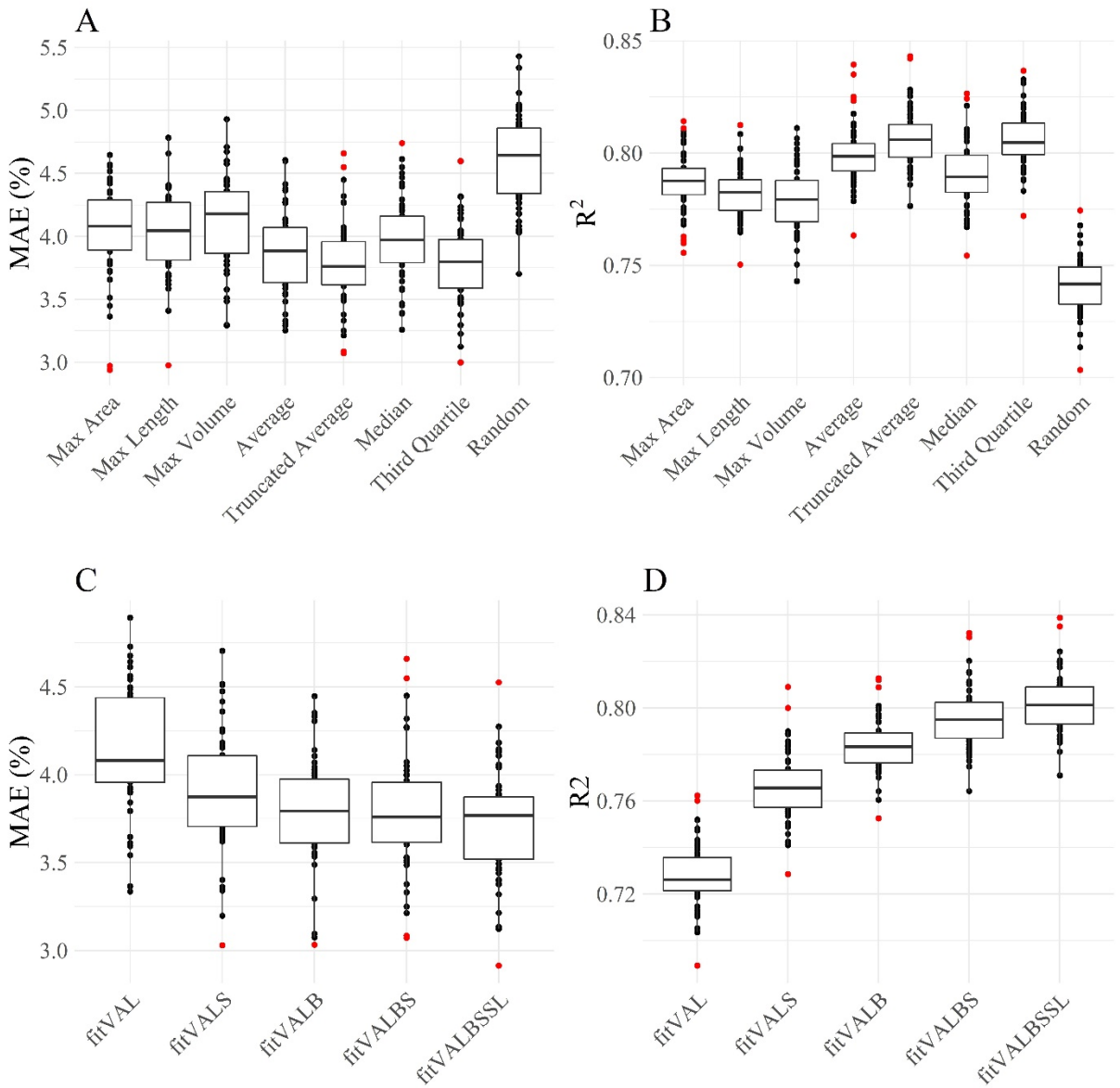

Figure 3.8 *Results for the different models across cross-validation for the dataset with only finishing pigs. The hinges represent the 1.5 interval interquartile (approximately 95% confidence interval) and red dots outlier results. Mean absolute error (MAE) as a percentage of the average body weight and coefficient of determination (R^2) for comparison between data reduction (**A** and **B**) and for inclusion of variables in the model (**C** and **D**). **fitVAL** stands for a model with volume, area, and length; **fitVALS** includes also shape descriptors; **fitVALB** includes also body measures; **fitVALBS** includes both shape descriptors and body measures; **fitVALBSSL** includes shape descriptors, body measures, sex, and line information.*

random image for each animal to 0.92 for the truncated mean and third quartile datasets (Figure 3.7-B). However, after removing the data from nursery animals, the R^2 ranged from 0.74 for the randomly selected image dataset to 0.81 for the truncated mean and third quartile (Figure 3.8-B). However, this increase of R^2 when including the data from nursery animals did not translate to lower MAE (Figure 3.7-A and Figure 3.8-A) since the median MAE for both scenarios is within the same intervals.

The importance of using only body volume, area, and length, and the inclusion of some of the remaining body measurements (heights and widths), shape descriptors, and sex and line effects were also assessed (Figure 3.8-C and 3.7-D). The selected widths were W1, W2, W5, W6 and W8, and the selected heights were H3, H4, H9, H10, and H11, whereas for the shape descriptors the remaining variables were eccentricity and the polar Fourier descriptors of order 1, 2, 4, 10, 11, 12 and 13. These variables were selected based on the variable appearance results for the datasets with and without nursery animals. There was a clear improvement of including body measurements or shape descriptors compared with using only the information of volume, area, and length. However, there was no visible improvement when including both shape and body measurements or the effect of sex and line after accounting for all body measurement variables. A graph with predicted BW vs. measured BW on the third quartile dataset for the finishing pigs is presented for the model fitVALBS, which included both body measurements and shape descriptors (Figure 3.9). The correlation between the predicted

and measured BW was 0.88, and the MAE was 3.6%, which corresponds to 4.36 kg for animals with an average BW of 120 kg.

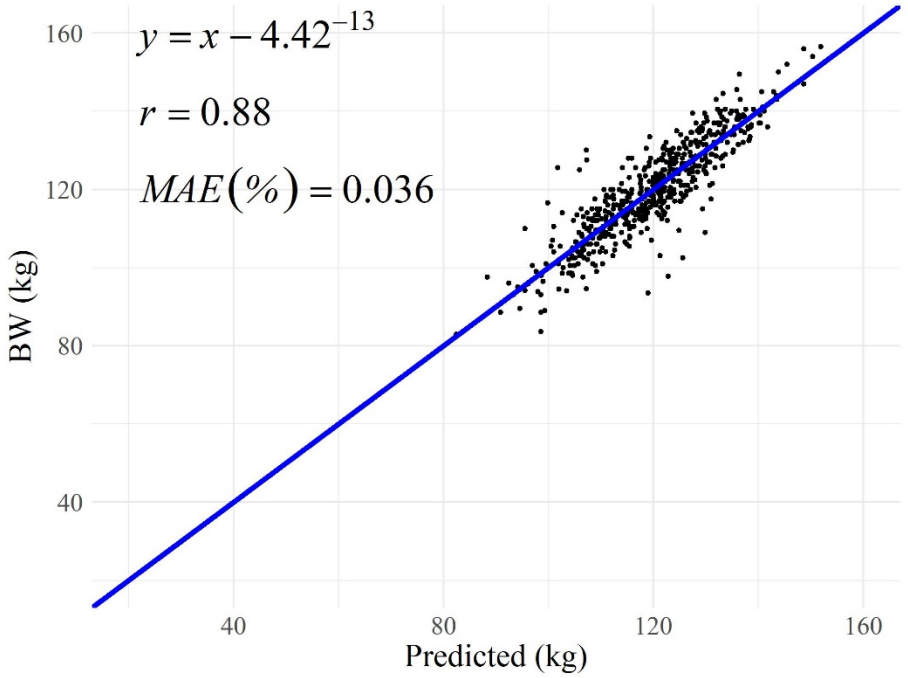

Figure 3.9 *Regression of body weight (BW) on the predicted BW for the model including volume, area, length, selected widths (W1, W2, W5, W6, and W8), heights (H3, H4, H9, H10, and H11), eccentricity and the polar Fourier descriptors of order 1, 2, 4, 10, 11, 12, and 13, the correlation (r) between selected and predicted and the mean absolute error (MAE) in percentage for the third quantile dataset including only finishing pigs.*

3.5 Discussion

The current study presents a framework for real-time video segmentation and extraction of image features for prediction of BW of pigs in commercial farms that does not require any manual selection or manipulation of the images. To evaluate the effect of including animals of different age classes, analyses with the full dataset (including younger animals) and with a subset with only finishing pigs were performed. The inclusion of data from nursery pigs increased the range of BW to a minimum of 27 kg and a maximum of 156 kg, whereas for the dataset without nursery pigs the range was from 83

to 156 kg. As a consequence, the R^2 from the models including younger animals were higher, with an average of 0.92 (Figure 3.7), whereas for models without nursery pigs the average R^2 was of 0.80 (Figure 3.8). This higher R^2 estimate is actually an artifact due to the increase in the data range, with a consequent increase on the correlation between the predictor's body measures and BW followed by an increase of the R^2 statistics. However, this increase in R^2 was not reflected in the MAE. Actually, the best predictions for the test dataset in the CV including only finishing pigs had an MAE of 3% while the best models when including the nursery pigs had an MAE of 3.5%. Therefore, in a linear model setting, the inclusion of the wide range of the data across the growth period is forcing a model that is actually predicting the average weight at a given age. This may not be optimal for commercial pig operations, in which predictions of BW within age classes may be necessary.

In previous work, Kongsro (2014) demonstrated the application of 3D cameras for the estimation of pig body weight with a reported residual mean square error (RMSE) of 4.8% (3.38 kg) and an R^2 of 0.99 for 71 animals with BW ranging from 30 to 140 kg. Recently, Condotta et al. (2018) performed a similar study, with more individuals (234 pigs), and reported a standard error of 3.13 kg and R^2 of 0.99 for animals of 10 to 125 kg. The findings in the current study for the dataset with nursery animals were within the previously reported range. In another study, Pezzuolo et al. (2018) evaluated not only the possibility of using a Kinect sensor to predict BW but also for measurement of biometric traits for pigs within 6 and 46 kg. In their study, the biometric traits measured from the images had high correlations with the manual measurements (from 0.77 for back height to 0.93 for heart girth). The main contribution of our work is the full automation of the

whole process of image selection, processing, and analysis. To the best of the researchers' knowledge, the only other work in the literature that presented results from a fully autonomous image analysis for prediction of pig body weight was Kashiha et al. (2014). In their work, however, the images were obtained using a surveillance camera (grayscale image) positioned at a specific distance from the pen floor. Images from standard digital cameras have the disadvantage of producing images that are strongly influenced by environmental conditions. These light differences plus differences in background conditions make it difficult for practical applications of these cameras with their image segmentation algorithms. Therefore, the algorithms developed by Kashiha et al. (2014) are optimum for white animals on a dark floor, and thus, they are not robust for implementation in commercial farms with many different backgrounds, illumination conditions, and animal coat colors. On the other hand, the depth image is invariant with regard to the background color and suffers only from extreme light conditions that can increase the noise at the pixel level. Also, with direct measurement of the distance from the sensor to the floor, it is possible to retrieve information of animal measurements in the metric scale (i.e. transform from pixels to centimeters) without the need for setting the CVS at a fixed known distance.

Even though the approach of reducing the dataset presented was efficient in removing images of animals with a bad position, achieving similar prediction error to previous works, there is still variation within the set of images for a given animal after the removal of outlier images (Figure 3.10). Possible reasons for this variation across images from the same animal may be due to the following: 1) There is natural variation in the images retrieved by the algorithm since the animals are moving; 2) Some of the retrieved

images may be non-optimum, which can result in a cascade of reduced quality in segmentation and feature extraction; and 3) The internal variability of the Kinect depth sensor that has a resolution of 2 mm and a variation on the order of 6mm (Yang et al., 2015). Thus, the image processing can still be further improved. Another possibility is of outlier animals in the group of finishing pigs. This is because within animals with similar BW, there are animals with a leaner body composition and others with more fat deposition. In fact, several factors can influence fat deposition, such as nutrition and genetics. For example, Reyer et al. (2017) studied the genetics of body fat mass of pigs from the same breeding line and observed that animals took from 124 to 167 d to reach 110 kg of BW, with a lean mass ranging from 56 to 66%, and subcutaneous fat thickness from 4.5 to 14 mm. These differences in body composition may imply differences in shape and biometric measures for animals of the same weight class, which will increase the error in predicting BW with the current methods. Therefore, additional studies are needed in order to further improve the prediction quality of BW for finishing pigs.

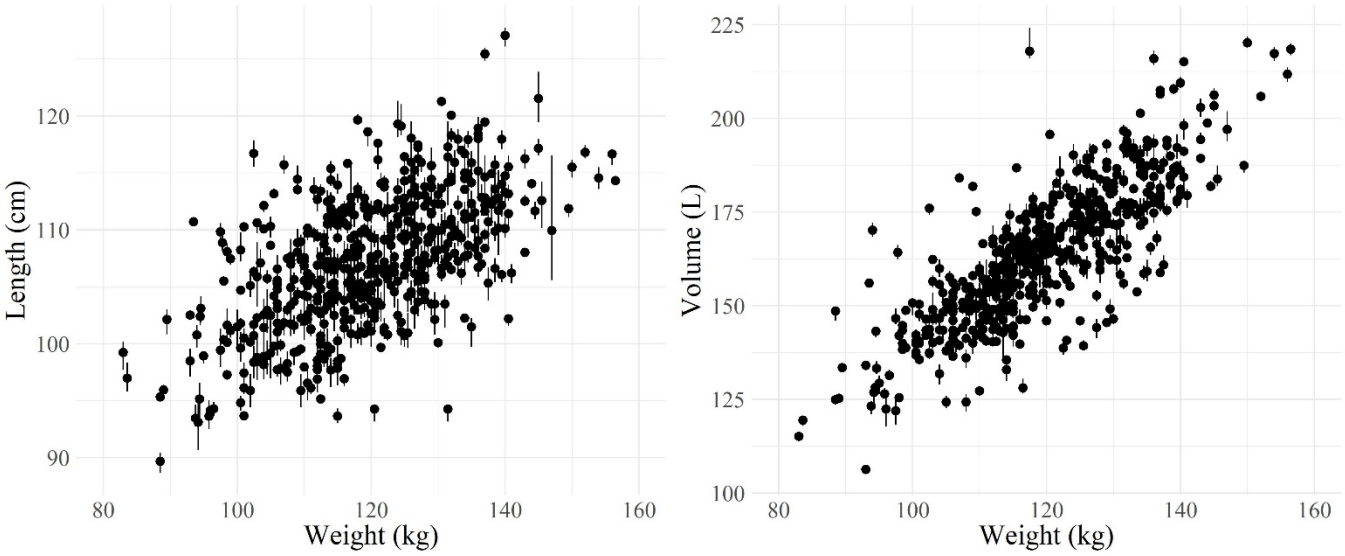

Figure 3.10 *Averages (dots) and interquartile interval (vertical lines) for body length (the most variable biometric trait) and volume (the least variable biometric trait) measured across the frames for each animal vs. body weight for finishing animals.*

In the present application, the possibility of using shape descriptors in the linear model in order to account for variation in animal position and segmentation was evaluated. Body measurements as height and length among others were used previously as predictor variables for BW since they have a positive correlation with BW. But, shape descriptors as Fourier transforms are a more general group of image features that are used in order to classify images of different objects and shape recognition (Bowman et al., 2000; Zhang and Lu, 2002), and are not necessarily linearly correlated with BW. The incorporation of shape descriptors to the linear model increased the predictive ability in comparison to the model containing only volume, area, and length information. However, the inclusion of shape descriptors to the model containing all selected body measurements did not show a significant improvement in MAE, even though there was an increase in R^2 (Figure 3.8 C and D). Thus, other approaches need to be devised in order to account for this systematic variation due to animal positioning. One possibility is to use more rigorous thresholds on animal position. However, this can reduce the number of images retrieved per animal. Another approach would be to include a classification step after the image segmentation or the use of nonlinear models that can account for these small differences in animal posture. This may be a more efficient approach and may be able to account also for differences in body conformation across animals of the same age group.

In this study, an entire pen of pigs was moved to a weighing and managing area. Another practice, more common in commercial farms, is to position a scale inside the pig pen and weigh just a fraction of the animals. This allows the farmer to have an estimate of the pen average weight and variability. Ultimately, the presented approach should be

adapted to be used with a camera inside a pig pen, so all pigs have weight predictions, without the need of moving or stressing the animals. This is a topic for further research and development so that a full CVS application can be made available to the pig industry.

Currently, with the increasing size of farms and the need for increasing productivity, farmers are faced with the challenge of efficiently managing more animals. This raises the demand for automated tools for monitoring animal growth and health. Thus, CVS applications have been proposed in pig production for tracking of animal activity level (Kashiha et al., 2013; Lind et al., 2005; Tillett et al., 1997), behavior (Lao et al., 2016; Nasirahmadi et al., 2015; Viazzi et al., 2014), and gait and locomotion (Kongsro, 2013; Stavrakakis et al., 2015). The methods presented here can be integrated with other applications, the most direct being the prediction of gait and walking kinematics (Stavrakakis et al., 2015) that used similar sensors. Also, the proposed CVS can be easily modified to function in conjunction with animal tracking algorithms for the collection of data in barn conditions without the need for human interaction. This provides the possibility for several measurements of pig body weight and biometric measurements across time. With this information, precise prediction of the growth curve for each animal can be obtained, allowing the farmer to better define marketing strategies (Cadéro et al., 2018; Conte et al., 2012; Tokach and Henry, 2008). This would be interesting since within the traditional management system, it is not practical to weigh pigs several times due to the labor involved and intensive stress caused to the animals. Moreover, the integration with other methods as described above can provide the farmer with detailed information not only on a pen (average) level but also for each pig. A successful application should be integrated also with an intuitive user interface that allows farmers to remotely access

the information obtained. By having this information easily accessible, farmers will be able to have a live report of the farm performance status for timely management decisions, e.g., to better pinpoint animals with health issues.

3.6 Conclusion

A fully automated system using the 3D camera for online extraction of body measurements of interest and prediction of pig BW was presented. The main differences between the current approach and the previous studies are the absence of manual manipulation of the images and the inclusion of measurement of shape descriptors. Therefore, the presented approach moves towards the implementation of a CVS for the acquisition of biometric traits and BW on commercial farms. The current implementation is not able to identify animal ID and still requires the need for moving the animals. For a successful commercial application, there is still the need to incorporate means to correctly track individual ID in a large group of animals without the need for human interaction with the pigs.

The truncated average and truncated median on the third quartile were the dataset reduction strategies that achieved lower MAE. Incorporation of sex and line on the model did not improve predictions compared with the models that accounted for shape descriptors and body measurements. Overall, this result shows that it is possible to achieve high predictive accuracy with only the information retrieved from the images.

3.7 References

Atherton, T.J., Kerbyson, D.J., 1999. Size invariant circle detection. Image Vis. Comput. 17, 795–803. https://doi.org/10.1016/S0262-8856(98)00160-7

Bowman, E.T., Soga, K., Drummond, T.W., 2000. Particle Shape Characterisation using Fourier Analysis. CUED/D-Soils/TR315 20.

Bradley, D., Roth, G., 2007. Adaptive Thresholding using the Integral Image. J. Graph. Tools 12, 13–21. https://doi.org/10.1080/2151237X.2007.10129236

Burger, W., Burge, M.J., 2016. Digital image processing : An Algorithmic Introduction Using Java, 2nd ed. Springer London, London. https://doi.org/10.1007/978-1-4471-6684-9

Cadéro, A., Aubry, A., Brossard, L., Dourmad, J.Y., Salaün, Y., Garcia-Launay, F., 2018. Modelling interactions between farmer practices and fattening pig performances with an individual-based model. Animal 12, 1277–1286. https://doi.org/10.1017/S1751731117002920

Condotta, I.C.F.S., Brown-Brandl, T.M., Silva-Miranda, K.O., Stinn, J.P., 2018. Evaluation of a depth sensor for mass estimation of growing and finishing pigs. Biosyst. Eng. https://doi.org/10.1016/j.biosystemseng.2018.03.002

Conte, S., Lawlor, P.G., O'Connell, N., Boyle, L.A., 2012. Effect of split marketing on the welfare, performance, and carcass traits of finishing pigs. J. Anim. Sci. 90, 373–380. https://doi.org/10.2527/jas.2010-3453

de Lange, C.F.M., Marty, B.J., Birkett, S., Morel, P., Szkotnicki, B., 2001. Application of pig growth models in commercial pork production. Can. J. Anim. Sci. 81, 1–8. https://doi.org/10.4141/A00-006

FASS, 2010. Guide for the Care and Use of Agricultural Animals in Research and Teaching, 3rd Editio. ed. Federation of Animal Science Societies, Champaign, IL.

Faucitano, L., Goumon, S., 2018. Transport of pigs to slaughter and associated handling, in: Advances in Pig Welfare. Woodhead Publishing, Cambridge, MA - USA, pp. 261–293. https://doi.org/10.1016/B978-0-08-101012-9.00009-5

Gous, R., Morris, T.R., Fisher, C., 2006. Mechanistic modelling in pig and poultry production. CABI.

Grandin, T., Shivley, C., 2015. How Farm Animals React and Perceive Stressful Situations Such As Handling, Restraint, and Transport. Animals 5, 1233–1251. https://doi.org/10.3390/ani5040409

Hough, P.V.C., 1962. METHOD AND MEANS FOR RECOGNIZING COMPLEX PATTERNS. US 3069654.

Kashiha, M.A., Bahr, C., Ott, S., Moons, C.P.H., Niewold, T.A., Ödberg, F.O., Berckmans, D., 2014. Automatic weight estimation of individual pigs using image analysis. Comput. Electron. Agric. 107, 38–44. https://doi.org/10.1016/j.compag.2014.06.003

Kashiha, M.A., Bahr, C., Ott, S., Moons, C.P.H., Niewold, T.A., Tuyttens, F., Berckmans, D., 2013. Automatic Monitoring of Pig Activity Using Image Analysis. Springer, Cham, pp. 555–563. https://doi.org/10.1007/978-3-319-02895-8_50

Kongsro, J., 2014. Estimation of pig weight using a Microsoft Kinect prototype imaging system. Comput. Electron. Agric. 109, 32–35. https://doi.org/10.1016/j.compag.2014.08.008

Kongsro, J., 2013. Development of a computer vision system to monitor pig locomotion. Open J. Anim. Sci. 3, 254–260. https://doi.org/10.4236/ojas.2013.33038

Lao, F., Brown-Brandl, T., Stinn, J.P., Liu, K., Teng, G., Xin, H., 2016. Automatic recognition of lactating sow behaviors through depth image processing. Comput. Electron. Agric. 125, 56–62. https://doi.org/10.1016/j.compag.2016.04.026

Lind, N.M., Vinther, M., Hemmingsen, R.P., Hansen, A.K., 2005. Validation of a digital video tracking system for recording pig locomotor behaviour. J. Neurosci. Methods 143, 123–132. https://doi.org/10.1016/j.jneumeth.2004.09.019

Microsoft, 2014. Kinect SDK for Windows.

Nasirahmadi, A., Richter, U., Hensel, O., Edwards, S., Sturm, B., 2015. Using machine vision for investigation of changes in pig group lying patterns. Comput. Electron. Agric. 119, 184–190. https://doi.org/10.1016/j.compag.2015.10.023

Pezzuolo, A., Guarino, M., Sartori, L., González, L.A., Marinello, F., 2018. On-barn pig weight estimation based on body measurements by a Kinect v1 depth camera. Comput. Electron. Agric. 148, 29–36. https://doi.org/10.1016/J.COMPAG.2018.03.003

R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Reyer, H., Varley, P.F., Murani, E., Ponsuksili, S., Wimmers, K., 2017. Genetics of body fat mass and related traits in a pig population selected for leanness. Sci. Rep. 7, 9118. https://doi.org/10.1038/s41598-017-08961-4

Stavrakakis, S., Li, W., Guy, J.H., Morgan, G., Ushaw, G., Johnson, G.R., Edwards, S.A., 2015. Validity of the Microsoft Kinect sensor for assessment of normal walking patterns in pigs. Comput. Electron. Agric. 117, 1–7. https://doi.org/10.1016/j.compag.2015.07.003

Szelisk, R., 2011. Computer Vision : Algorithms and Applications. Springer-Verlag, London. https://doi.org/10.1007/978-1-84882-935-0

Terven, J.R., Córdova-Esparza, D.M., 2016. Kin2. A Kinect 2 toolbox for MATLAB. Sci. Comput. Program. 130, 97–106. https://doi.org/10.1016/j.scico.2016.05.009

The MathWorks, 2017. MATLAB Release 2017b.

Tillett, R.D., Onyango, C.M., Marchant, J.A., 1997. Using model-based image processing to track animal movements. Comput. Electron. Agric. 17, 249–261. https://doi.org/10.1016/S0168-1699(96)01308-7

Tokach, M., Henry, S., 2008. When, Where and Why: Marketing Decisions Based on Packer Matrixes and Pig Flow, in: Swine Profitability Conference. Manhattan, KS, pp. 49–62.

Venables, W.N., Ripley, B.D., 2002. Modern applied statistics with S. Springer-Verlag New York. https://doi.org/10.1007/978-0-387-21706-2

Viazzi, S., Bahr, C., Van Hertem, T., Schlageter-Tello, A., Romanini, C.E.B., Halachmi, I., Lokhorst, C., Berckmans, D., 2014. Comparison of a three-dimensional and two-dimensional camera system for automated measurement of back posture in dairy cows. Comput. Electron. Agric. 100, 139–147. https://doi.org/10.1016/j.compag.2013.11.005

Yang, L., Zhang, L., Dong, H., Alelaiwi, A., Saddik, A. El, 2015. Evaluating and Improving the Depth Accuracy of Kinect for Windows v2. IEEE Sens. J. 15, 4275–4285. https://doi.org/10.1109/JSEN.2015.2416651

Zhang, D., Lu, G., 2002. Shape-based image retrieval using generic Fourier descriptor. Signal Process. Image Commun. 17, 825–848.

Chapter 4: Comparison of multiple linear, partial least squares, elastic network and artificial neural network models for prediction of pig body weight and fat and muscle depths from 3D images

4.1 Abstract

Computer vision systems (CVS) have been shown to be a powerful tool for non-invasive measurement of live pig body weight. With advances in precision farming, it is now possible to evaluate the growth performance of individual pigs more accurately. However, important traits such as muscle and fat deposition can still be evaluated only via ultrasound, computed tomography, or dual x-ray absorptiometry. Therefore, the objectives of this study were to develop CVS for prediction of live body weight (BW), muscle depth (MD), and back fat (BF) from top view 3D images of finishing pigs and to compare different modeling approaches such as traditional multiple linear regression, partial least squares, and machine learning approaches, such as elastic networks, artificial neural networks, and deep learning. A dataset containing approximately 12,000 images from 618 finishing pigs (average BW of 120 kg) was divided into training (80% of the data) and testing sets. Several image features, such as volume, area, length, widths, heights, polar image descriptors, and polar Fourier transforms were extracted from the images and used as predictor variables in the model. In addition, the raw 3D images were also evaluated as potential model inputs. A deep learning model achieved the lowest mean absolute errors (mean absolute error / average) on the test dataset of 3.47 kg

(2.89%), 3.45 mm (5.33%) and 0.95 mm (15.7%) on the prediction of BW, MD, and BF, respectively. Nonetheless, the predictions of BF were quite similar to those obtained with the other approaches evaluated. In conclusion, it was demonstrated that it is possible to successfully predict BW, MD, and BF via CVS on a fully automated setting using 3D images from farm conditions. Moreover, with the use of a deep learning model, the raw 3D image can be used as model input without the need of prior image processing and still get relatively high prediction accuracy.

4.2 Introduction

It is forecasted that the human population will be about 8.6 billion people worldwide by 2030 and that the overall consumption of meat will increase substantially, following current trends, as low-income countries have more than tripled their overall meat consumption (FAO, 2018). To meet this increasing demand, farmers are faced with the challenge of increasing productivity while reducing the use of human-edible crops and the waste produced (FAO, 2017). In this scenario, a concern is to maintain or improve animal welfare and optimum production indices while increasing the ratio of animal product produced per unit of labor. Thus, there is a need for changes in the livestock production systems in order to increase the efficiency in monitoring animal welfare and productivity.

Precision livestock farming (PLF) proposes to address this challenge via the use of sensors, such as activity trackers and cameras (Benjamin and Yik, 2019; Berckmans, 2017). From the many possible applications, the use of computer vision systems (CVS) is particularly appealing since they provide non-invasive monitoring of animal behavior

and high-throughput measurements of traits of interest. In pig production, the literature presents applications for tracking of pig behavior, identification of leg and back disorders and measurement of pig weight, among others (Fernandes et al., 2019; Kashiha et al., 2013; Maselyne et al., 2017; Stavrakakis et al., 2014). Automatic measurement of weight on live pigs is viewed as a critical management tool for precision adjustment of feed quantity and for scheduling of pig sales. In this sense, more important than the live weight would be the possibility to measure lean muscle and fat deposition in live pigs. Currently, these traits are eventually measured via ultrasound (US), dual-energy x-ray absorptiometry (DXA) or computed tomography (CT) (Carabús et al., 2016; Scholz et al., 2015). Even though DXA and CT produce the most accurate measurements, these methods are expensive and require the animal to be anesthetized, making their use impractical in commercial operations. On the other hand, ultrasound scanning technologies are portable and do not require pigs to be anesthetized. Nonetheless, they are generally less accurate than DXA or CT. CVS has also been proposed for the estimation of body composition of live animals, with the biggest advantage of allowing indirect measurement of the body surface of the pigs. However, there are few works on the use of CVS for measurement of muscle and fat composition in live pigs, using either features extracted from 2D images or reconstruction of the pig body shape via computer stereo vision (Doeschl-Wilson et al., 2005; Wu et al., 2004).

In this context, the objectives of this study were: (1) to develop a CVS for prediction of live body weight (BW), muscle depth (MD), and back fat (BF) from top view 3D images of finishing pigs; and (2) to compare different predictive modeling approaches including

multiple linear regression, partial least squares, and machine learning approaches, such as elastic networks, artificial neural networks, and deep learning image encoder.

4.3 Material and Methods

The data sets of video recordings and measurements of BW, MD, and BF were supplied by the Pig Improvement Company (PIC, a Genus company, Hendersonville, TN). The data collection followed rigorous animal-handling procedures that are in compliance with federal and institutional regulations regarding proper animal care practices (FASS, 2010).

4.3.1 Animals and Data Acquisition

A total of 618 finishing pigs were used, including boars and gilts from 3 different PIC commercial lines with average BW of 120 kg (ranging from 83 to 153 kg) and an average age of 151 ± 3 days, all on the same multiplier farm and raised under standard commercial settings. Pigs from the same pen were moved to a management area for evaluation, where they had their identification tag read using an RFID (radio frequency identification) sensor, followed by measurements of BW, MD, BF, and videos recorded for each pig. For measurement of BW an EziWeigh5i (Tru-Test, Mineral Wells, TX, USA) electronic scale which has a measured standard error of ±1% of the weight load was used. Video recordings from each pen were acquired using a Microsoft Kinect V2 (Microsoft, Redmond, WA, USA), which is a multi-sensor device equipped with RGB digital camera (1920 × 1080 pixels of resolution), time of flight depth sensor (resolution of

512 × 424 pixels), infrared camera (resolution of 512 × 424 pixels) and a microphone array. MD (mm) and BF (mm) were measured using an Aloka SSD 500 ultrasound device (Hitachi-Aloka, Tokyo, Japan) equipped with a 3.5 MHz, 12 cm linear probe. The measurements were collected by placing the probe perpendicular to the loin at the 10th intercostal space. For consistency of measurements, at least three consecutive readings were obtained before storing the values. The data acquisition was performed in the course of 3 months from February to April of 2016, and all data were collected by the same team of trained professionals.

4.3.2 Image Processing and Feature Extraction

From each video recording, images from the pigs were extracted along with their respective features. This process was completely automated with the video recordings emulated on a virtual Kinect camera via the Kinect for Windows SDK v2.0 (Microsoft, 2014) and a CVS for processing the video feed initialized on an independent MATLAB routine, with the connection between both instances performed using a custom encapsulation of the Kinect for Windows SDK on a MATLAB MEX following the directions in the Kin2 toolbox (Terven and Córdova-Esparza, 2016). The CVS used for video processing and extraction of suitable images and their respective image features corresponds to a library of custom codes developed by the authors in MATLAB (Release 2017b) (The MathWorks, 2017). This CVS is a combination of image processing and segmentation steps for selection of frames in which there is a pig well positioned, and subsequent removal from the image of the background, and head, and tail of the animal.

The algorithm then saves the original selected image along with the segmented one and respective extracted features (Fernandes et al., 2019).

In the current study, the features extracted from the images were: 1) body measurements from the pigs, including apparent volume (V), surface area (A), length (L), heights (H) and widths (W) collected at 11 equidistant points along the animal's back, and eccentricity (E); 2) 360 equidistant measurements from the polar shape contour of the top view image; and 3) the corresponding 360 Fourier descriptor features from the same polar shape contour. The body measurements extracted from the depth images were converted to the values on the metric scale by using the intrinsic focal length (*f*) from the Kinect depth camera. The f was used for estimation of the image magnification factor (m) in each pixel as $m = f/d$, where d is the distance in mm from the point to the camera plane. From m, the pixel area (a_p) in mm was calculated as $a_p = {}^{1}/_{m}$, and the total area of the pig calculated as the sum of a_p. Moreover, the volume of a pixel is then $v_p = a_p * h_p$, and so the pig volume was calculated as the sum of pixel volumes, where h_p is the difference between the floor distance to the camera and the distance from the pixel on the back of the pig to the camera. The pig eccentricity was estimated as the ratio between the foci and the major axis of the ellipsis that has the same major and minor axis as the pig area. The polar shape descriptors (PSD) used were measured as the distance from the centroid (x_c, y_c) of the pig to points on its boundary contour as $PSD_t = \sqrt{(x_c - x_t)^2 + (y_c - y_t)^2}$, where the point (x_t, y_t) is the coordinate of the pig boundary at the degree t. The polar Fourier descriptors (PFD) used were the fast Fourier transform from the polar transform of the pig contour (Ostermeier et al., 2001; Zhang and Lu, 2002). For robustness, both

shape descriptors were extracted from the rotated and centralized segmented pig image with animal head and tail removed (Figure 4.1).

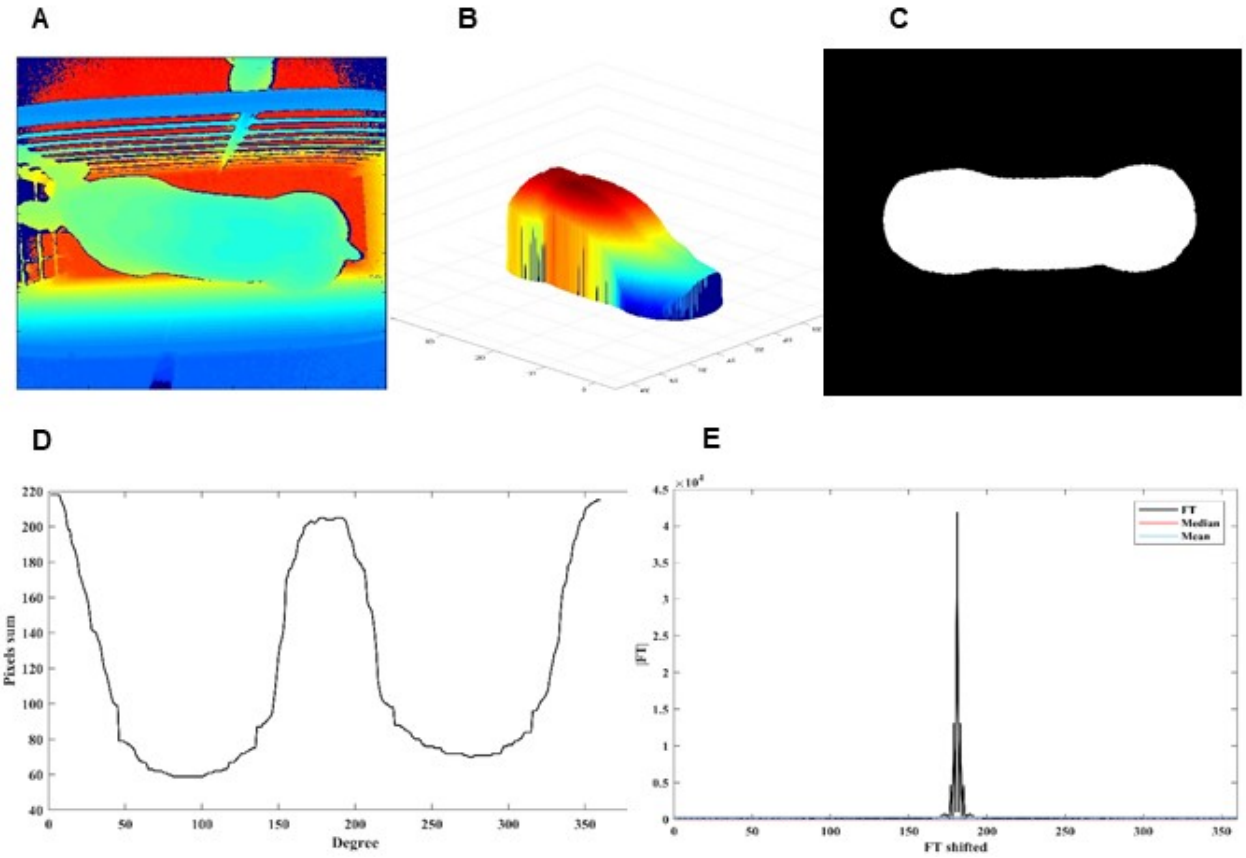

Figure 4.1 *Image processing pipeline. A) Original depth image. B) Extracted pig surface. C) Rotated and centralized binary mask of the extracted pig surface. D) Distance in pixels from the pig centroid to the contour by degree, the polar shape descriptors. E) The polar Fourier transform from D.*

4.3.3 Statistical Analysis

Different approaches for prediction of BW, MD, and BF were evaluated. For this, the dataset was divided into training (80% of the pigs) and validation sets. The criteria used to compare these models were the mean absolute error (MAE), root mean square error (RMSE), and the squared predictive correlation (R^2). The models evaluated can be

divided into two groups: 1) models that used as predictor variables the metrics extracted from the image processing step (these included Multiple Linear Regression, Partial Least Squares, Elastic Network Regression, and Artificial Neural Networks) and 2) Deep Learning Image encoder models, which used the raw 3D images as model input. A description of each of these modeling approaches and of the specific predictive model techniques is provided below.

4.3.4 Models using Metrics from Images as Predictor Variables

A total of 746 metrics were extracted from the images, including 26 biometrics, such as volume (V), area (A), length (L), eccentricity (E), 11 heights (H), and 11 widths (W), and 360 PSD and 360 PFD image descriptors. These features plus information on pig sex and genetic line were used as predictor variables on the models that follow. Since each pig had several images acquired but only one value of BW, MD, and BF, the truncated median at the third quartile of each predictor variable was calculated as described in Fernandes et al. (2019). Thus, the final dataset used in this section had just one value (the truncated median) per pig for each of the features extracted from the images.

4.3.4.1 Multiple Linear Regression

The multiple linear regression models (ML) were developed in R (R Core Team, 2017) using the MASS package (Venables and Ripley, 2002). To avoid overfitting a stepwise approach was used for variable selection using the *stepAIC* function from the MASS package, with the Bayesian information criteria (BIC) for model selection. The

multiple linear regression models were fitted on the training set and had their predictive performance evaluated on the validation set.

4.3.4.2 Partial Least Squares

The partial least squares (PLS) models were developed using the *pls* package in R (Mevik and Wehrens, 2007). For this analysis, a 10-fold cross-validation (CV) on the training data was performed for selection of the most significant latent variables. The selection of the best set of latent variables was performed using the randomization strategy with a 5% significance level (Mevik and Wehrens, 2007). Starting with the model that reached the absolute optimum at the CV, this strategy removes latent variables until there is no significant deterioration of the model performance at the specified significance level.

4.3.4.3 Elastic Network Regression

The elastic networks (EN) were developed using the *glmnet* package in R (Friedman et al., 2010). In the current work, the EN evaluated were multiple linear regressions with a mixture of the L1 (lasso) and L2 (ridge) penalties, in which the mixture of the two regularizations is controlled by the hyperparameter α that ranges from 0 (a full ridge penalty) to 1 (lasso penalty). Six different values of alpha were tested, from 0 to 1 in intervals of 0.2. Apart from α, the EN has an additional parameter λ that controls the shrinkage of the regression coefficients. Thus, for each value of α, a 10-fold CV was performed on the training set for estimation of the value of λ that gives the minimum CV error.

4.3.4.4 Artificial Neural Networks

Artificial neural networks (ANN) were developed using the H2O platform for machine learning that has a package in R (LeDell et al., 2019). ANN, also known as a multi-layer perceptron, is a class of machine learning algorithms with an arbitrary number of features that have the purpose of learning non-linear combinations of the input variables in order to predict a given output (Murphy, 2012). In ANN models there are many possible parameters to be manipulated and tuned, such as the number of layers and nodes in each layer, activation function, loss function, regularization parameters, such as L1 and L2 penalizations, and dropout rate, among others. Therefore, one of the biggest challenges in fitting ANN is to define the ANN architecture. In the present study the ANN models evaluated differed in: 1) number of hidden layers and nodes, from a single layer perceptron to at most 3 hidden layers with a number of nodes in each layer ranging from 5 to 100; 2) the activation function, that could be a rectified linear unit (ReLU) or max-out (Goodfellow et al., 2013), both with a constraint on the squared sum of the incoming weights that could be from 5 to 100 and dropout rate of input and hidden layer nodes from 20 to 80% of the nodes in each layer; 3) the loss functions evaluated were a Gaussian loss (in this case, the mean squared error) or the Huber loss as a robust alternative (Hastie et al., 2009); 4) the inclusion of the L1 and L2 regularizations with their specific shrinkage parameters; and 5) the learning rate and time decay of the AdaDelta adaptive learning algorithm.

The search on the model space for possible architectures was conducted using a random discrete search on 500 candidate models, from which the 10 best candidate

models were later fine-tuned. All the models were fitted on the training set and subsequently had their predictive performance evaluated on the validation set.

4.3.5 Deep Learning Image Encoder Models

The deep learning image encoder models (DL) were developed in Python (version 3.7) using the TensorFlow machine learning library (Abadi et al., 2015). These models had as inputs an array with 2 channels, one containing the depth image of a pig and the second with the focal length map from the camera that generated the depth image. The basic architecture of the image encoder models was a multi-layer perceptron with encoder block similar to state of the art segmentation networks (Poudel et al., 2019), composed of a convolutional block, followed by a max-pooling layer with a 2 by 2 window and strider of the same size. The convolutional block was composed by a convolutional layer with a 3 by 3 window followed by batch normalization and ReLU activation. The last encoder block is followed by two fully connected (FC) layers with L1 and L2 regularization, dropout and leaky ReLU activation function. The encoder models tested diverged mainly on the size of the input image, which ranged from 0.20 to 0.5, and a number of nodes on FC layers, which ranged from 4 to 64.

4.4 Results

4.4.1 Descriptive Statistics of Image Features

Figure 4.2 presents the correlations between biometric traits and the three traits of interest (BW, BF, and MD). Overall, BW was found to be highly correlated with volume and area, and positively correlated with BF and MD and most of the biometric traits, except for eccentricity which showed a low negative correlation with BW. Muscle depth presented a low positive correlation with most of the other traits, with the highest values being with BW, V, A and widths W1, W2, W9, and W10. BF also presented low positive correlations with most of the other traits, with the highest value with BW. In addition, the

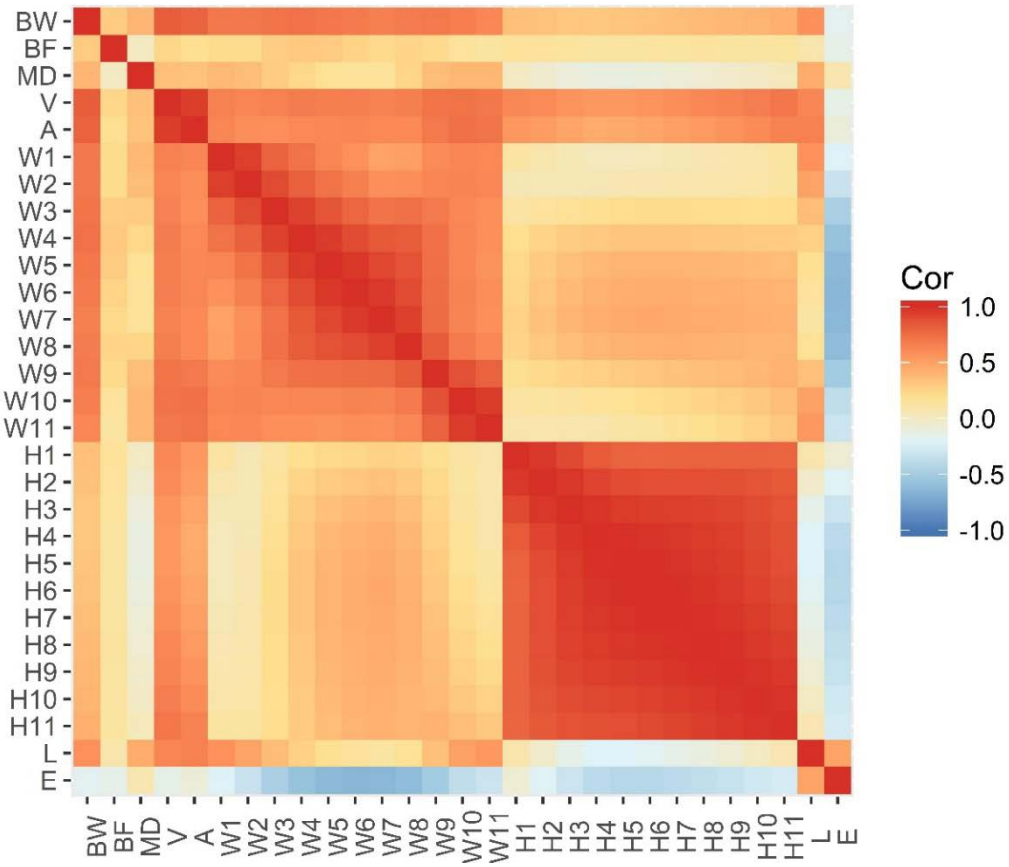

Figure 4.2 *Correlations between body weight (BW), back fat (BF) muscle depth (MD) and the extracted biometric traits, volume (V), area (A), widths (W), heights (H), length (L), eccentricity (E).*

biometric traits were mostly positively correlated with each other; notably there was a higher correlation within widths and within heights. The exception was for eccentricity, which presented mostly negative correlations with the other traits, except with MD and length. Eccentricity was notably negatively correlated with widths W4, W5, W6, and W7.

The correlations between BW, BF, MD, and the PSD are presented in Figure 4.3-A. PSD presented higher correlations with BW than with MD or BF. Also, the correlation between PSD and BW or MD followed the same pattern with higher correlations with PSD from 1 to 25, 160 to 200 and from 335 to 360. However, for the correlations between PSD with BF, this pattern was shifted. Overall, there was a high correlation between neighboring PSD values. It is worth noticing that there were two big blocks of high correlated PSD, one from 30 to 160 degrees, and another from 210 to 335 degrees. There were also two small blocks of high positively correlated PSD, one from 165 to 209 degrees and another from 345 going back to 20 degrees. Lastly, there was a low negative correlation between PSDs going from 30 to 160 with the opposite block, going from 220 to 330 degrees. Regarding the correlations between the PFD and BW, MD and BF, there

was no clear pattern, with few PFD having a high positive correlation with the traits of interest.

4.4.2 Model Performance for Prediction of BW, MD, and BF

Table 4.1 shows the results of MAE, RMSE, and R^2 for prediction of BW, MD, and BF for the various modeling strategies evaluated. For BW there was virtually no difference between the performance of the different strategies except for DL, which presented better predictive performance with an MAE of 3.47 kg, which is on average 25% smaller than the MAE achieved with the other modeling strategies, and an R^2 of 0.80, the second largest being 0.69 for EN. For MD, the DL approach again achieved the best performance,

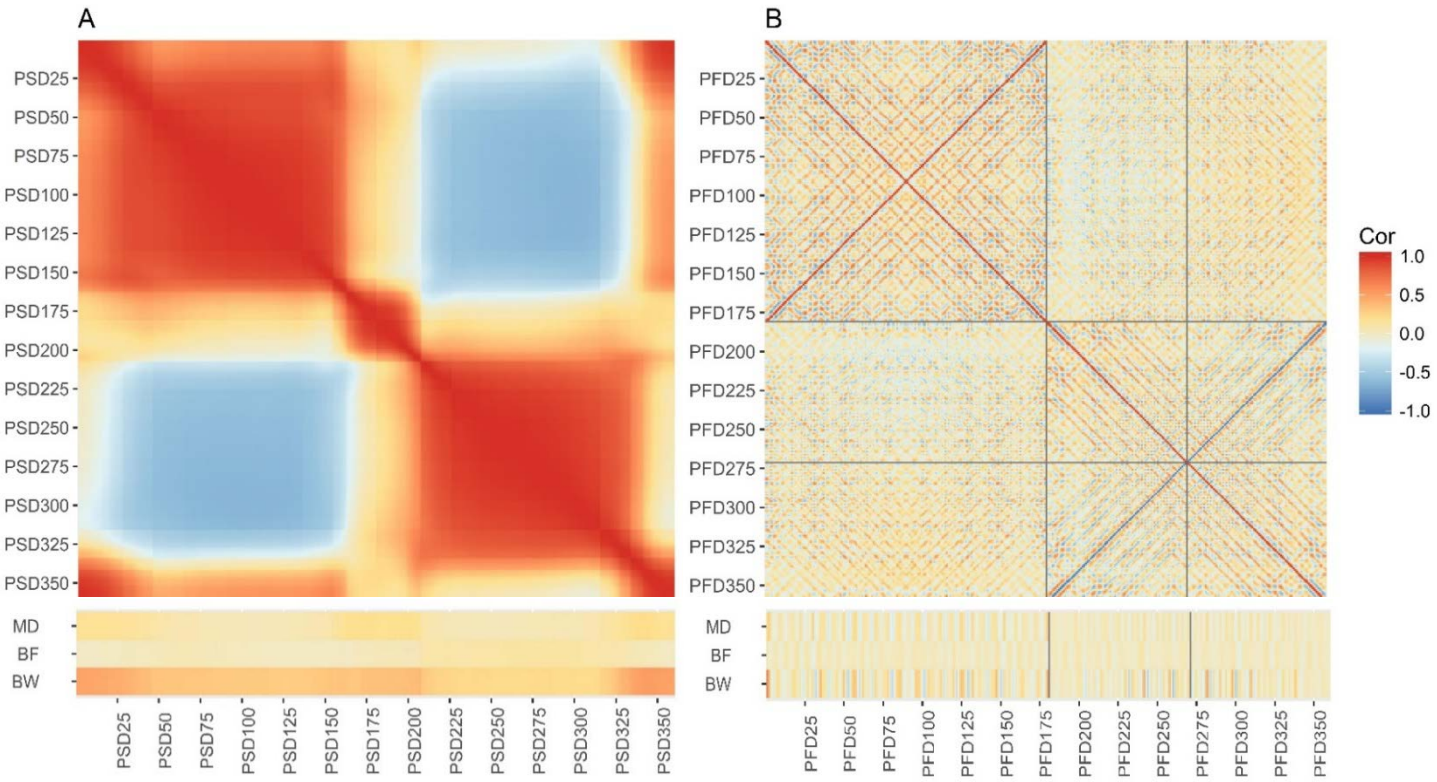

Figure 4.3 *A) Correlations between body weight (BW), back fat (BF) muscle depth (MD) and the polar shape descriptors (PSD) extracted as the distances of the pig contour to the centroid at each degree. B) Correlations between BW, BF, MD and the polar Fourier descriptors (PFD) transform of the PSD.*

with an MAE of 3.45 mm which is on average 30% smaller than the MAE achieved by the other models. However, the R^2 for MD was of 0.47, which is lower than the R^2 achieved for BW. Regarding BF, even though DL achieved the lowest values of MAE and RMSE, it was close to the performance of the other models evaluated. Also, the R^2 was of only 0.22 for DL, and the best value of R^2 was of 0.26 for the LM.

Table 4.1 *Estimated mean absolute error (MAE), root mean square error (RMSE) and squared predictive correlation (R^2) for body weight (BW), muscle depth (MD) and back fat (BF) on the test dataset for the best multiple linear regression model (LM), partial least squares (PLS), elastic network (EN), artificial neural network (ANN), and deep learning image encoder (DL) evaluated*

Predicted trait	Model	MAE	RMSE	R^2
BW (kg)	LM	4.36	5.89	0.68
	PLS	4.49	5.95	0.66
	EN	4.24	5.70	0.69
	ANN	4.30	5.70	0.68
	DL	3.47	4.71	0.80
MD (mm)	LM	4.56	5.64	0.27
	PLS	4.37	5.48	0.32
	EN	4.45	5.51	0.31
	ANN	4.63	6.02	0.20
	DL	3.45	4.62	0.47
BF (mm)	LM	1.15	1.42	0.26
	PLS	1.21	1.48	0.18
	EN	1.16	1.43	0.25
	ANN	1.19	1.48	0.18
	DL	0.95	1.21	0.22

4.5 Discussion

The current study attempted to improve predictions of BW of pigs from past works by applying statistical and machine learning approaches that have not been used before.

Previous experimental results on the use of image features from 3D images of growing pigs for prediction of BW have already shown the potentials of CVS applications (Condotta et al., 2018; Fernandes et al., 2019; Kongsro, 2014). However, those applications required some level of human processing of the images or manual measurement of the image features of interest. To the best of the authors' knowledge, there are only two other works that applied a fully automated approach for extraction of features from images and, subsequently, prediction of BW in pigs (Fernandes et al., 2019; Kashiha et al., 2014). From these previous CVS applications, only Fernandes et al. (2019) presented results for finishing pigs (151 ± 2.6 days of age, and BW of 120 ± 12.4 kg). The other available works presented predictions for growing animals from different ages, with BW ranging from 15 to 50 kg (Kashiha et al., 2014) and from 20 to 120 kg (Condotta et al., 2018; Kongsro, 2014). The predictive performance achieved in the current work was similar to what was observed in previous studies, with an error of approximately 2-4% of the animals' BW. From the approaches evaluated, DL achieved the highest performance in the test dataset with a model that had as input an image of 128 x 106 pixels, 4 encoder layers, and 2 fully connected layers of 16 and 8 nodes in each. The main advantages of DL over the other techniques are that there is no need for splitting the task of prediction of BW from images in many steps, such as image processing, feature extraction and later prediction of the trait of interest. Therefore, the model automatically searches for the best encoder features and weights that reduce the model loss (e.g., the mean squared error for Gaussian models).

Pig carcass composition is of great importance for farmers since packing plants tend to reward them according to lean muscle percentage (Engel et al., 2012). Moreover,

information on individual pig growth and muscle and fat deposition can be used by farmers to improve feeding strategies as feed costs account for approximately 60 to 70% of total pig production. In addition, there is an increasing concern on fattening pigs emissions of phosphorus and nitrogen (Cadéro et al., 2018), and a well-balanced feed can reduce such costs and emissions. Imaging techniques such as visual image analysis, US, DXA, and CT have been traditionally used in the evaluation of pig carcass composition and also to evaluate live pigs in research conditions with CT and DXA presenting overall higher accuracy than other techniques (Carabús et al., 2016). Previous study evaluating US, CT, and slaughter measurements of MD and BF showed that US measurements have a correlation of 0.6 and 0.56 with the carcass measurements of MD and BF, respectively, while CT showed correlations of 0.48 to 0.67 for fat and 0.91 to 0.94 for lean meat (Font-i-Furnols et al., 2015; Lucas et al., 2017). However, there is little information in the literature regarding the uses of CVS for evaluation of live pig lean muscle and fat composition. In the only other study found that evaluated the use visual image analysis for prediction of pig muscle and body fat composition (Doeschl-Wilson et al., 2005), the predictions of whole body fat or lean muscle composition using only image features achieved an adjusted R^2 of 0.20 on average. In the current study, the MAE on MD was 3.45 mm (corresponding to 5.33%) of the average MD, which was of 65.04 ±6.21 mm, and the R^2 on the test dataset was of 0.47 for the selected DL model (Table 4.1). On the other hand, the MAE on BF was 0.95 mm, which is around 15.7% of the average BF that was of 6.06 ± 1.63 mm, with an R^2 on the test dataset of 0.22 for the DL model. Also, this predictive performance was similar to the selected LM model (Table 4.1) that included as predictor variables volume, W5, W6, pig line, and PSD 355.

BF and MD were virtually uncorrelated, but both were positively correlated with BW. Therefore, one could question if the predictive approaches developed in this study are better than using only BW for prediction of MD or BF. By fitting a linear regression of MD on BW we found a MAE of 4.84 mm, RMSE of 6.04 and R^2 of 0.17 on the test dataset, which are worse than the predictive accuracy of all the approaches evaluated. Similarly, for BF we found higher values of MAE (1.29 mm) and RMSE (1.53 mm), and lower R^2 (0.14) than what was observed for the predictive approaches evaluated (Table 4.1). Thus, even though BW is an important trait in the prediction of MD and BF the approaches developed in this study are in the end accounting for more information than only sing BW as predictor variable.

In regard to the correlations between BF and the traits extracted from the images, and between MD and the traits extracted from the images it is interesting to notice that they were generally greatly different. This trend was observed for L, widths and the PSD, since BF presented higher correlations with W4 to W6 while MD had higher correlations with W1, W2, W9 and W10 (Figure 4.1). Similarly, BF showed a higher correlation with the central PSDs measured around 225 to 300 degrees, while BW and MD had higher correlations with PSDs 1 to25 and 335 to 360 degrees (Figure 4.2). Interestingly, the correlations observed are in accordance with results from Peñagaricano et al. (2015), who investigated phenotypic/genetic causal networks of muscle and fat deposition in pigs. In their study, an antagonistic relationship existed between fat and muscle deposition, with back fat having a negative causal effect on loin depth and the presence of a strong genetic marker at chromosome 6.

In the current study, single trait models were developed for each trait of interest, thus any field application of the current models would need to deploy at least 3 different predictive models, one for each trait (BW, BF, and MD). Since there is shared information among the three traits, a possible future research direction would be to evaluate the predictive performance of models for joint prediction of these three traits. One such approach is multiple trait models, that could possibly improve the predictive accuracy of MD and BF by aggregating the information of the correlated traits that are been jointly predicted. Another possible development would be the use of other methods for measurement of muscle and fat composition instead of US. Alternatives could be DXA and CT as these methods have proven to provide more accurate measurements of muscle and fat composition that could be used to develop predictive models for a CVS based on 3D images.

4.6 Conclusion

In conclusion, a DL model using the raw depth images provided higher prediction accuracy for BW and MD than the other methods evaluated. It was demonstrated that is possible to predict MD and BF via CVS on a fully automated setting using 3D images from farm conditions, without the need of preprocessing images in steps such as image segmentation and feature extraction. Therefore, the methods developed here can be extended to commercial pig farms. However, prediction of BF via the use of CVS was more challenging. As such, additional research in this area is warranted,

especially regarding the improvement of the direct measurement of fat composition from live pigs to be used as a target trait for the development of predictive models.

4.7 References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., Google Research, 2015. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems [WWW Document]. URL https://www.tensorflow.org/about/bib (accessed 6.20.19).

Benjamin, M., Yik, S., 2019. Precision Livestock Farming in Swine Welfare: A Review for Swine Practitioners. Animals 9, 133. https://doi.org/10.3390/ani9040133

Berckmans, D., 2017. General introduction to precision livestock farming. Anim. Front. 7, 6. https://doi.org/10.2527/af.2017.0102

Cadéro, A., Aubry, A., Brossard, L., Dourmad, J.Y., Salaün, Y., Garcia-Launay, F., 2018. Modelling interactions between farmer practices and fattening pig performances with an individual-based model. Animal 12, 1277–1286. https://doi.org/10.1017/S1751731117002920

Carabús, A., Gispert, M., Font-i-Furnols, M., 2016. Imaging technologies to study the composition of live pigs: A review. Spanish J. Agric. Res. 14, e06R01. https://doi.org/10.5424/sjar/2016143-8439

Condotta, I.C.F.S., Brown-Brandl, T.M., Silva-Miranda, K.O., Stinn, J.P., 2018. Evaluation of a

depth sensor for mass estimation of growing and finishing pigs. Biosyst. Eng. https://doi.org/10.1016/j.biosystemseng.2018.03.002

Doeschl-Wilson, A.B., Green, D.M., Fisher, A. V, Carroll, S.M., Schofield, C.P., Whittemore, C.T., 2005. The relationship between body dimensions of living pigs and their carcass composition. Meat Sci. 70, 229–240. https://doi.org/10.1016/j.meatsci.2005.01.010

Engel, B., Lambooij, E., Buist, W.G., Vereijken, P., 2012. Lean meat prediction with HGP, CGM and CSB-Image-Meater, with prediction accuracy evaluated for different proportions of gilts, boars and castrated boars in the pig population. Meat Sci. 90, 338–344. https://doi.org/10.1016/J.MEATSCI.2011.07.020

FAO, 2018. Shaping the future of livestock sustainably, responsibly, efficiently, in: 10th Global Forum for Food and Agriculture. FAO, Berlin, p. 20.

FAO, 2017. Livestock solutions for climate change. https://doi.org/I8098EN/2/12.17

Fernandes, A.F.A., Dórea, J.R.R., Fitzgerald, R., Herring, W., 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. https://doi.org/10.1093/jas/sky418

Font-i-Furnols, M., Carabús, A., Pomar, C., Gispert, M., 2015. Estimation of carcass composition and cut composition from computed tomography images of live growing pigs of different genotypes. animal 9, 166–178. https://doi.org/10.1017/S1751731114002237

Friedman, J., Hastie, T., Tibshirani, R., 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. J. Stat. Softw. 33, 1–22.

Goodfellow, I.J., Warde-Farley, D., Mirza, M., Courville, A., Bengio, Y., 2013. Maxout Networks.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning, Second Edi. ed. Springer New York, New York, NY. https://doi.org/10.1007/978-0-387-84858-7

Kashiha, M.A., Bahr, C., Haredasht, S.A., Ott, S., Moons, C.P.H., Niewold, T.A., Ödberg, F.O., Berckmans, D., 2013. The automatic monitoring of pigs water use by cameras. Comput.

Electron. Agric. 90, 164–169. https://doi.org/10.1016/j.compag.2012.09.015

Kashiha, M.A., Bahr, C., Ott, S., Moons, C.P.H., Niewold, T.A., Ödberg, F.O., Berckmans, D., 2014. Automatic weight estimation of individual pigs using image analysis. Comput. Electron. Agric. 107, 38–44. https://doi.org/10.1016/j.compag.2014.06.003

Kongsro, J., 2014. Estimation of pig weight using a Microsoft Kinect prototype imaging system. Comput. Electron. Agric. 109, 32–35. https://doi.org/10.1016/j.compag.2014.08.008

LeDell, E., Gill, N., Aiello, S., Fu, A., Candel, A., Click, C., Kraljevic, T., Nykodym, T., Aboyoun, P., Kurka, M., Malohlava, M., Rehak, L., Eckstrand, E., Hill, B., Vidrio, S., Jadhawani, S., Wang, A., Peck, R., Wong, W., Gorecki, J., Dowle, M., Tang, Y., DiPerna, L., 2019. R Interface for “H2O” [WWW Document]. CRAN. URL https://cran.r-project.org/web/packages/h2o/h2o.pdf (accessed 6.17.19).

Lucas, D., Brun, A., Gispert, M., Carabús, A., Soler, J., Tibau, J., Font-i-Furnols, M., 2017. Relationship between pig carcass characteristics measured in live pigs or carcasses with Piglog, Fat-o-Meat’er and computed tomography. Livest. Sci. 197, 88–95. https://doi.org/10.1016/J.LIVSCI.2017.01.010

Maselyne, J., Van Nuffel, A., Briene, P., Vangeyte, J., De Ketelaere, B., Millet, S., Van den Hof, J., Maes, D., Saeys, W., 2017. Online warning systems for individual fattening pigs based on their feeding pattern. Biosyst. Eng. 1–14. https://doi.org/10.1016/j.biosystemseng.2017.08.006

Mevik, B.-H., Wehrens, R., 2007. The pls Package: Principal Component and Partial Least Squares Regression in R. J. Stat. Softw. 18, 1–23. https://doi.org/10.18637/jss.v018.i02

Microsoft, 2014. Kinect SDK for Windows.

Murphy, K.P., 2012. Machine Learning: A Probabilistic Perspective. MIT Press, Cambridge, Massachusetts.

Ostermeier, G.C., Sargeant, G.A., Yandell, B.S., Parrish, J.J., 2001. Measurement of Bovine Sperm Nuclear Shape Using Fourier Harmonic Amplitudes. J. Androl. 22, 584–594.

https://doi.org/10.1002/J.1939-4640.2001.TB02218.X

Peñagaricano, F., Valente, B.D., Steibel, J.P., Bates, R.O., Ernst, C.W., Khatib, H., Rosa, G.J., 2015. Exploring causal networks underlying fat deposition and muscularity in pigs through the integration of phenotypic, genotypic and transcriptomic data. BMC Syst. Biol. 9, 58. https://doi.org/10.1186/s12918-015-0207-6

Poudel, R.P.K., Liwicki, S., Cipolla, R., 2019. Fast-SCNN: Fast Semantic Segmentation Network. arxiv 1–9.

R Core Team, 2017. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing. Vienna, Austria.

Scholz, A.M., Bünger, L., Kongsro, J., Baulain, U., Mitchell, A.D., 2015. Non-invasive methods for the determination of body and carcass composition in livestock: dual-energy X-ray absorptiometry, computed tomography, magnetic resonance imaging and ultrasound: invited review. animal 9, 1250–1264. https://doi.org/10.1017/S1751731115000336

Stavrakakis, S., Guy, J.H., Warlow, O.M.E., Johnson, G.R., Edwards, S.A., 2014. Walking kinematics of growing pigs associated with differences in musculoskeletal conformation, subjective gait score and osteochondrosis. Livest. Sci. 165, 104–113. https://doi.org/10.1016/j.livsci.2014.04.008

Terven, J.R., Córdova-Esparza, D.M., 2016. Kin2. A Kinect 2 toolbox for MATLAB. Sci. Comput. Program. 130, 97–106. https://doi.org/10.1016/j.scico.2016.05.009

The MathWorks, 2017. MATLAB Release 2017b.

Venables, W.N., Ripley, B.D., 2002. Modern applied statistics with S. Springer-Verlag New York. https://doi.org/10.1007/978-0-387-21706-2

Wu, J., Tillett, R., McFarlane, N., Ju, X., Siebert, J.P., Schofield, P., 2004. Extracting the three-dimensional shape of live pigs using stereo photogrammetry. Comput. Electron. Agric. 44, 203–222. https://doi.org/10.1016/J.COMPAG.2004.05.003

Zhang, D., Lu, G., 2002. Shape-based image retrieval using generic Fourier descriptor. Signal

Process. Image Commun. 17, 825–848.

Chapter 5: Concluding Remarks

This dissertation focused on the development and application of computer vision systems in animal production, more specifically for monitoring growth and carcass-related traits in live Nile tilapia and pigs. The questions addressed revolve around: 1) applications using images from standard digital cameras and 3D sensors for identification and segmentation of animals from images captured under farm conditions (i.e., uncontrolled and diverse settings of illumination and background conditions), 2) the relationship between production traits of interest, such as body and carcass weights and lean meat and fat content with features and biometric traits measured from images, and 3) the performance of different modeling strategies for the prediction of such traits of interest.

Monitoring of production traits, such as animal growth and carcass-related traits, is of great importance to any animal production system. Animal growth performance is directly related to the overall health, and the ability to identify animals that are underperforming can enable farmers to intervene sooner and more effectively. Moreover, the ability to monitor animal growth is desired in order to best determine nutritional strategies and to prevent both under and overfeeding. This is of great importance because feed is approximately 70% of animal production costs, and incorrect feeding strategies lead to reduced profitability. Apart from animal production, constantly monitoring of animal growth and carcass traits is also important for breeding purposes, since multiple measurements can increase the prediction accuracy of an animal's breeding value. However, manual measurement of any growth trait currently requires the

direct interaction with the animals, which is stressful to the animals and costly to the farmers and, thus, prohibitive to be done on a routine basis.

Computer vision systems (CVS) are an interesting alternative for continuous measurement of animal growth since a CVS does not require direct interaction with the animals. Moreover, a system capable of measuring or predicting traits of interest automatically without much input from the user is desirable. It is known that biometric traits, such as body length and height, are highly correlated with body weight. In Chapter 2, instead of manual measurements of biometric traits or using images from fish with trimmed fins as in previous works, a deep learning approach for fish body segmentation was used. We found that it was possible to accurately segment fish body from fins and background on digital images. Moreover, fish body area was found to be highly correlated with body weight and carcass weight and thus sufficient for accurate prediction of both.

In pig farms, for both production and breeding purposes, body weight is generally measured at most 3 times, i.e., on transfer from sow to nursery, from nursery to finishing farm, and ultimately from finishing to packing plant. In order to increase the number of measurements, a computer vision system based on images from surveillance cameras has been proposed before. However, such a system has the drawback that it cannot be used in situations where the animal skin color is similar to the floor color (e.g., pigs with dark skin on a dark floor). Alternatively, depth sensors have been proposed for the extraction of pig volume and prediction of body weight. The depth sensors are useful for extraction of pigs from background since the resultant image is invariant to animal or background color. However, every implementation previously proposed required some sort of manual processing of the images, which is undesirable for daily use of this system

in pig farms. In Chapter 3, a computer vision system for automatic identification of pigs on depth images and extraction of biometric traits and image features was proposed. The devised automated system presented prediction accuracies similar to what has been previously achieved by manually processing the images.

Lastly, Chapter 4 of this dissertation addressed the problem of improving the prediction accuracy of live pig body weight and the possibility of using depth images to predict muscle and fat deposition of finishing pigs. Muscle and fat deposition is also of great importance for pig farmers, as many packing plants reward according to the muscle and fat ratio. Moreover, along with pig weight, these two traits are an indicator of correct nutritional management and animal growth. It was found that it is possible to achieve better prediction accuracy for both pig body weight and muscle depth using a deep learning model that has the raw depth image as input. These results are interesting since there is no need for image processing and extraction of image features, as the image processing and prediction of the desired traits are jointly performed on a single step.

Ultimately, it is important to highlight that the methods developed in this dissertation are only part of what may be a full implementation of computer vision systems for animal production. Any implementation of the methods developed here would also require the development of a software interface that is appropriate and useful for daily farm operations. Even though providing tools to the farmers is the ultimate goal of any scientific work in animal sciences, it was not the focus of this dissertation. Overall, results from this dissertation support the potential of successful application of computer vision systems to aid in more efficient and optimized animal production.