

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Bovine segmentation and anatomical keypoint identification

Boru Galma Elema

WORKING VERSION



Master in Computer Vision

Supervisor: Prof. Pedro Miguel Cravalho

Second Supervisor: Dr. Ana Santos (PhD)

July 10, 2026

Resumo

As medições morfométricas precisas desempenham um papel essencial na gestão pecuária, apoiando a monitorização do crescimento, o melhoramento genético, a avaliação da saúde animal e a pecuária de precisão. As técnicas convencionais de medição baseiam-se em procedimentos manuais, que são trabalhosos, demorados e suscetíveis a erros humanos, tornando a sua aplicação difícil em explorações pecuárias de grande escala. Os avanços na visão computacional oferecem oportunidades para automatizar estas medições através da análise de imagens sem contacto, embora desafios como fundos complexos, variações de iluminação e alterações na postura dos animais continuem a afetar a fiabilidade dos métodos existentes.

Esta dissertação propõe uma estrutura baseada em visão computacional para a segmentação de bovinos e identificação de pontos anatómicos utilizando imagens RGB monoculares adquiridas em condições reais de exploração pecuária. A estrutura proposta integra calibração de escala baseada em laser, segmentação do animal e identificação de pontos anatómicos numa única cadeia de processamento para a estimativa automática de medições morfométricas. A deteção dos pontos laser fornece uma referência fiável para converter medições obtidas na imagem em dimensões reais, enquanto o Segment Anything Model (SAM) é utilizado para isolar o corpo do animal. Posteriormente, os pontos anatómicos são identificados através de uma abordagem heurística que combina a estimação da pose com a análise geométrica baseada em contornos, permitindo estimar as principais características morfométricas dos bovinos.

A estrutura proposta foi avaliada utilizando um conjunto de imagens de bovinos adquirido em condições reais de exploração pecuária. Os resultados experimentais demonstram que a estrutura proporciona uma localização fiável dos pontos laser, uma segmentação precisa do animal e uma estimativa consistente das medições morfométricas em diferentes condições ambientais. De um modo geral, a metodologia proposta constitui uma solução de baixo custo, não invasiva e escalável para a análise morfométrica automática de bovinos, contribuindo para o avanço da pecuária de precisão e estabelecendo uma base para futuros sistemas de estimação do peso vivo e de monitorização inteligente de animais.

Abstract

Accurate morphometric measurements play an essential role in livestock management by supporting growth monitoring, breeding, health assessment, and precision livestock farming. Conventional measurement techniques rely on manual procedures that are labor-intensive, time-consuming, and susceptible to human error, making them difficult to apply in large-scale farming environments. Advances in computer vision provide opportunities to automate these measurements through non-contact image analysis, although challenges such as complex backgrounds, varying illumination, and changes in animal pose continue to affect the reliability of existing methods.

This dissertation proposes a computer vision framework for bovine segmentation and anatomical keypoint identification using monocular RGB images acquired under real farm conditions. The proposed framework integrates laser-based scale calibration, animal segmentation, and anatomical keypoint identification into a unified processing pipeline for automatic morphometric measurement. Laser point detection provides a reliable reference for converting image measurements into real-world dimensions, while the Segment Anything Model (SAM) is employed to isolate the animal body. Anatomical landmarks are then identified using a heuristic approach that combines pose estimation with contour-based geometric analysis to estimate key bovine morphometric traits.

The proposed framework was evaluated using a cattle image dataset collected under practical farm conditions. The experimental results demonstrate that the framework provides reliable laser localization, accurate animal segmentation, and consistent estimation of morphometric measurements under diverse environmental conditions. Overall, the proposed methodology offers a low-cost, non-invasive, and scalable solution for automated cattle morphometric analysis, contributing to the advancement of precision livestock farming and providing a foundation for future body weight estimation and intelligent livestock monitoring systems.

UN Sustainable Development Goals

The United Nations Sustainable Development Goals (SDGs) provide a global framework for addressing social, economic, and environmental challenges while promoting sustainable development. This dissertation contributes to sustainable livestock farming through the development of computer vision techniques for automated bovine segmentation, anatomical keypoint identification, and morphometric measurement extraction. By enabling non-invasive and automated livestock monitoring, the proposed framework supports improved productivity, resource efficiency, and data-driven decision-making in animal production systems.

The specific Sustainable Development Goals addressed by this research are presented in Table 1.

Table 1: Contribution of the dissertation to the United Nations Sustainable Development Goals.

SDG	Target	Contribution	Performance Indicators and Metrics
2	2.4	The proposed framework supports sustainable livestock production by enabling automated and non-invasive monitoring of cattle growth and body morphology. Accurate morphometric measurements can improve animal management practices, feeding strategies, and production efficiency, contributing to sustainable agricultural systems and food security.	Morphometric measurement accuracy, segmentation performance, monitoring efficiency.
8	8.2	Automation of cattle morphometric measurements reduces labor-intensive manual procedures, improves operational efficiency, and minimizes human intervention during livestock monitoring. This contributes to productivity growth through technological innovation.	Reduction in manual workload, processing time, automation rate.
9	9.5	The dissertation applies advanced computer vision and artificial intelligence techniques, including laser-based calibration, image segmentation, pose estimation, and anatomical keypoint detection, to address challenges in livestock farming. This promotes innovation and digital transformation in agriculture.	Intersection over Union (IoU), Dice Score, Precision, Recall, F1-score, keypoint localization performance.
12	12.2	Accurate livestock monitoring enables more efficient resource utilization and data-driven decision-making regarding feeding, growth assessment, and production planning. This supports sustainable agricultural management and responsible production practices.	Measurement consistency, monitoring accuracy, resource management efficiency.

Acknowledgements

The completion of this dissertation would not have been possible without the support, encouragement, and generosity of many individuals and institutions, to whom I express my sincere gratitude.

My deepest appreciation goes to my supervisor, Professor Pedro Carvalho, for his invaluable guidance, continuous encouragement, insightful advice, and unwavering support throughout this research. His expertise, constructive feedback, and commitment to academic excellence have been fundamental to the successful completion of this dissertation and have greatly contributed to my academic and professional development.

Special thanks are extended to Dr. Ana Santos for her valuable support and for providing the cattle image dataset used in this research. I also gratefully acknowledge the project *Medida 1. “Inovação” do PDR 2020 — Programa de Desenvolvimento Rural do Continente — Grupo Operacional Preservar a Qualidade na Carne Arouquesa (PDR2020–101–031094)*, through which the dataset was made available. Their contribution was essential to the development and evaluation of the proposed framework.

My sincere appreciation is also extended to Camões – Instituto da Cooperação e da Língua, I.P., for awarding me the scholarship that enabled me to pursue my Master’s studies at the Faculty of Engineering of the University of Porto. This financial support provided the opportunity to undertake this research and successfully complete my studies.

I am also thankful to my colleagues and members of the research group for their valuable discussions, encouragement, and willingness to share their knowledge throughout this work.

To my beloved wife, I offer my deepest gratitude for her unconditional love, endless patience, unwavering encouragement, and constant belief in me. Her understanding, sacrifices, prayers, and support have been my greatest source of strength during every stage of this journey.

Finally, my heartfelt thanks go to my parents, family, and friends for their unconditional love, prayers, encouragement, and continuous support. Their confidence in me and their unwavering belief in my abilities have inspired me to persevere through every challenge. This achievement is as much theirs as mine.

Boru Galma Elema

Declaration of honour

I declare that this dissertation is my original work and has not previously been submitted or assessed in any other course or curricular unit, whether at this or any other institution. References to other authors (statements, ideas, thoughts), as well as the use of published works fully complies with the rules of attribution and are appropriately cited in the text and bibliography, in accordance with reference standards. I am aware that plagiarism and self-plagiarism constitute academic misconduct.

I further declare that I have not used Artificial Intelligence tools to produce any part(s) of this dissertation or, when such tools were used, their use and results are clearly indicated and were carefully reviewed for the detection of errors, inaccuracies, unfounded attributions, or other forms of content and argument bias, as well as for the determination of authorship.

I further declare that I have included studies carried out with multiple co-authors, having specified, with transparency and rigour, the contribution of each co-author, provided information on the possible existence of any other dissertation in which the same article has been included, reproduced the work in full with the permission of the journal in which it was published, and obtained the explicit written consent of all co-authors for the inclusion of the article.

Boru Galma Elema



Contents

1	Introduction	1
1.1	Context	1
1.2	Motivation	1
1.3	Objectives	2
1.4	Document Structure	3
2	Literature Review	4
2.1	Region of Interest Identification	4
2.2	Segmentation	5
2.2.1	U-Net	5
2.2.2	DeepLabV3+	6
2.2.3	Mask R-CNN	7
2.2.4	Feature Pyramid Network (FPN)	8
2.2.5	SAM	9
2.3	Laser-Based Reference Detection	11
2.4	Pose Estimation	11
2.4.1	Unsupervised Domain Adaptation (Synthetic to Real data)	12
2.4.2	ScarceNet	13
2.5	Anatomical keypoint identification	13
2.5.1	Early Vision-Based Methods	13
2.5.2	Deep Learning-Based Approaches	14
2.5.3	CNN-Based Keypoint Detection	15
2.5.4	Face and Behavior-Specific Keypoint Detection	16
2.6	Research Gap	17
2.7	Summary	18
3	Methodology	22
3.1	Data Preparation	22
3.1.1	Image Acquisition Setup	22
3.1.2	Data Annotation and Ground Truth	23
3.1.3	Preprocessing	24
3.1.4	Morphological Measurements Dataset	25
3.2	Proposed Framework	27
3.3	Framework Implementation	29
3.3.1	Laser Point Identification	29
3.3.2	Animal Body Segmentation	33
3.3.3	Morphometrics Measurement	34
3.4	Comparison with other models	39

3.4.1	Dataset and Experimental Setup	39
3.4.2	Preprocessing and Data Consistency	39
3.5	Evaluation Metrics	39
3.5.1	Laser Point Detection Evaluation Metrics	39
3.5.2	Segmentation Metrics	41
3.5.3	Measurement Metrics	41
4	Results	42
4.1	Laser Point Detection Results	42
4.1.1	Quantitative Results	42
4.1.2	Visual Results	43
4.2	Segmentation Results of the Proposed SAM-Based Approach	45
4.2.1	Qualitative Results	46
4.2.2	Results of Benchmark Models	47
4.2.3	Visual Results of Benchmark Models	48
4.2.4	Quantitative Results of Benchmark Models	50
4.3	Morphological Measurements Results	51
4.3.1	Visual Analysis of Morphological Measurements	51
4.3.2	Quantitative Morphometric Measurement Evaluation	55
5	Conclusions	57
5.1	Final Discussion	57
5.2	Future Work	58
	References	61

List of Figures

2.1	Architecture of U-Net for semantic segmentation. Extracted from: (Ronneberger, Fischer, and Brox 2015).	6
2.2	Encoder-Decoder Architecture of DeepLabV3+ Extracted from: (Feng et al. 2023)	7
2.3	The Mask R-CNN framework for instance segmentation Extracted from: (He et al. 2017b).	8
2.4	Architecture of the Segment Anything Model (SAM) showing the image encoder, prompt encoder, and mask decoder. Kirillov, Mintun, and Ravi (2023)	10
2.5	Overview of anatomical keypoint identification approaches in livestock, including classical vision-based methods, CNN-based detection, deep learning approaches, and behavior-specific keypoint analysis	17
3.1	Cattle images with different backgrounds from the used dataset	24
3.2	Standard anatomical measurements collected during dataset acquisition. The proposed framework estimates five representative morphometric traits: withers height (a_1), height at tail insertion (ANC), chest depth (a_4), body length (b_1), and rump length (b_2). The remaining measurements belong to the original morphometric protocol and are not considered in this study.	25
3.3	Overview of the proposed framework for cattle morphometric analysis.	28
3.4	Overall framework adopted in this dissertation. The methodology consists of three main stages: (1) laser point identification, (2) animal segmentation, and (3) anatomic keypoint identification and distance calculation. The detected laser points provide the scale reference, the segmentation stage isolates the animal body, and the identified anatomical landmarks are used to estimate morphometric measurements.	29
3.5	Reference laser colour samples extracted from the dataset.. . . .	30
3.6	Example of artefacts present in the binary mask after CIEDE2000 thresholding, including isolated noise pixels and fragmented laser regions.	32
3.7	Laser-guided prompt generation for animal segmentation. The detected laser markers are used to generate multiple equally spaced foreground prompts between the laser points. These prompts provide reliable guidance because they are located on the animal body.	34
3.8	Morphometric measurements considered in this study. a_1 : withers height; a_4 : chest depth; ANC: waist height; b_1 : body length; b_2 : rump length.	35
4.1	Binary mask obtained after CIEDE2000 thresholding. Candidate laser regions are separated from the background based on color similarity.	44
4.2	Refined binary mask after morphological processing. Noise pixels are reduced and the detected laser regions become more compact and connected.	44

4.3	Examples of final laser-point localization obtained through contour extraction and bounding-box detection. The detected laser regions are enclosed within bounding boxes and used as spatial references for scale calibration.	45
4.4	Representative cattle segmentation overlays generated using the Segment Anything Model (SAM). The highlighted regions correspond to the cattle body identified by the segmentation model. The examples demonstrate the segmentation outputs obtained under different acquisition conditions and background environments.	46
4.5	Representative binary segmentation masks generated using SAM. White pixels correspond to the detected cattle regions, while black pixels represent the background.	46
4.6	Representative segmentation overlays generated using SAM. The highlighted regions correspond to the cattle body identified by the segmentation model.	47
4.7	Representative segmentation results obtained using the U-Net model. From left to right: input image, SAM-generated pseudo-ground-truth mask, and predicted segmentation mask.	49
4.8	Representative segmentation results obtained using the FPN model. From left to right: input image, SAM-generated pseudo-ground-truth mask, and predicted segmentation mask.	49
4.9	Representative segmentation results obtained using the DeepLabV3+ model. From left to right: input image, SAM-generated pseudo-ground-truth mask, and predicted segmentation mask.	50
4.10	Representative contour extraction results generated from the cattle image dataset. The extracted contours preserved the geometric structure of the animal body while minimizing background interference, enabling reliable anatomical landmark localization for morphometric analysis under varying environmental conditions.	52
4.11	Representative examples of morphometric distance calculation obtained from the automatically identified anatomical keypoints. The generated measurement references illustrate the estimation of withers height (a_1), waist height (ANC), chest depth (a_4), body length (b_1), and rump length (b_2) from segmented cattle images.	53
4.12	Representative results of Stage 3 of the proposed framework: anatomical keypoint identification and distance calculation. The figure compares manually annotated reference landmarks with the anatomical keypoints automatically generated by the proposed method. The identified landmarks provide the geometric basis for estimating morphometric traits, including withers height (a_1), waist height (ANC), chest depth (a_4), body length (b_1), and rump length (b_2).	54
4.13	Representative outputs of the anatomical keypoint identification stage. The automatically generated landmarks provide a structured geometric representation of the cattle body and establish the reference points required for morphometric distance calculation. These landmarks are subsequently used to estimate morphometric traits, including withers height (a_1), waist height (ANC), chest depth (a_4), body length (b_1), and rump length (b_2).	54

List of Tables

1	Contribution of the dissertation to the United Nations Sustainable Development Goals.	iv
2.1	Comparison of Deep Learning-Based Segmentation Methods	10
2.2	Comparison of livestock image segmentation methods: results, limitations, and representative studies	19
2.3	Comparison of animal pose estimation approaches: performance and limitations .	20
2.4	Comparison of anatomical key point identification approaches in livestock vision systems	21
3.1	Definition of the anatomical measurements contained in the morphometric dataset.	26
3.2	Sample of the ground-truth dataset containing animal characteristics and morphometric measurements used in this study.	27
3.3	Mapping between anatomical keypoints and morphometric measurements.	37
4.1	Intersection over Union (IoU) results for laser point detection.	43
4.2	Euclidean distance error of laser-point center localization.	43
4.3	Quantitative performance comparison of benchmark segmentation models.	50
4.4	Ground-truth and predicted morphometric measurements for six representative cattle.	55
4.5	Overall quantitative performance of the proposed morphometric measurement framework evaluated on the complete dataset of 66 cattle.	56

List of Acronyms

A²-FPN	Attention Aggregation Feature Pyramid Network
APE	Animal Pose Estimation
ASPP	Atrous Spatial Pyramid Pooling
BiFPN	Bi-directional Feature Pyramid Network
CNN	Convolutional Neural Network
FPN	Feature Pyramid Network
GPU	Graphics Processing Unit
MDAM	Multi-scale Domain Adaptation Module
PANet	Path Aggregation Network
PPLBO	Progressive Pseudo-Label-Based Optimization
R-CNN	Region Based Convolutional Neural Network
ROI	Region of Interest
S²-FPN	Scale-Sequence Feature Pyramid Network
WS-CDA	Weakly- and Semi-Supervised Cross-Domain Adaptation

Chapter 1

Introduction

1.1 Context

In recent years, there has been an increasing demand for sustainable livestock products (Le Cozler et al. 2019; Quaresma et al. 2019). As a result, optimal animal management has become an important mean to improve production efficiency in the livestock industry and achieve economic benefits.

Traditionally, livestock morphological measurements have been obtained through visual assessment and manual tools, such as rulers and tape measures, conducted by trained personnel. While widely used, these methods are labor-intensive, time-consuming, and difficult to scale for continuous monitoring. Morphological measurements are nevertheless critical parameters for controlling livestock growth, tailoring feeding strategies, and maintaining or improving productivity.

To overcome the limitations of manual measurements, computer vision and image processing algorithms have been adopted to perform animal size measurements automatically in a non-human-dependent manner (Alexey Ruchay et al. 2021; Jessica Nye, L. M. Zingaretti, and Miguel Pérez-Enciso 2020). Additionally, pose estimation can assist not only in body-weight analysis, but also in livestock health monitoring, such as detecting lameness, which significantly affects reproduction, welfare, and productivity (K. Zhao et al. 2018).

1.2 Motivation

Traditional livestock measurement techniques, such as visual estimation and manual tools (e.g., rulers and tape measures), are commonly used by farmers and zoo-technicians to evaluate animal growth and body condition (Le Cozler et al. 2019). However, these methods are labor-intensive, time-consuming, and prone to human error. They are difficult to apply in large-scale or continuous operations, and they may expose farmers to risks due to aggressive animal behavior while also causing stress to livestock.

With increasing demand for sustainable livestock production, accurate morphological measurements have become essential for optimizing feeding strategies, improving productivity, and

supporting data-driven farm management. Furthermore, livestock monitoring extends beyond weight estimation; pose-estimation techniques enable automated health and welfare assessment, including the detection of lameness — a condition that significantly affects productivity and animal welfare (K. Zhao et al. 2018). Therefore, there is a growing need for automated, low-cost, and non-invasive systems capable of operating continuously without disturbing animals' natural behavior. Computer-vision-based solutions, including animal pose estimation and image-based morphological analysis, have emerged as promising alternatives, offering faster, safer, and more accurate livestock monitoring and body-measurement capabilities for precision farming (Alexey Ruchay et al. 2021; Jessica Nye, L. M. Zingaretti, and Miguel Pérez-Enciso 2020). Existing systems for animal measurement and body condition analysis—including optical detection devices (Pezzuolo, Sartori, and Marinello 2018) and imaging-based techniques for dairy cattle (Kuzuhara et al. 2015)—are often expensive, complex, or require long processing times. Such limitations hinder their widespread deployment in real farming environments, particularly where scalability, robustness, and low operational cost are critical.

In addition, many existing approaches focus on individual processing stages, such as segmentation, pose estimation, or measurement, in isolation. While effective under controlled conditions, this modular treatment can reduce robustness when applied in unconstrained farm settings, where variations in illumination, animal posture, background clutter, and occlusions are common.

These challenges motivate the development of an integrated computer vision framework that combines segmentation, anatomical landmark identification, and measurement within a single processing pipeline, designed specifically for practical deployment in livestock farming environments.

1.3 Objectives

The major objective of this research is to develop a computer vision-based pipeline for automated cattle segmentation, anatomical keypoint identification, and morphometric measurement using RGB images acquired under real livestock farm conditions. By integrating deep learning, image processing, and laser-guided calibration techniques, the proposed framework aims to provide a scalable, non-contact, and efficient solution for livestock body measurement and precision livestock monitoring.

To achieve this overall goal, the following specific objectives are defined:

- Develop and evaluate a robust laser point detection and calibration method for accurate localization of laser markers in cattle images.
- Benchmark and improve state-of-the-art cattle segmentation models for precise animal extraction from RGB images.
- Analyze and assess state-of-the-art bovine pose estimation approaches for livestock posture analysis and structural understanding.

- Develop an anatomical keypoint detection approach to identify relevant body landmarks from segmented cattle images.
- Convert image-based measurements into real-world dimensions using laser-guided calibration techniques.
- Integrate laser detection, animal segmentation, pose estimation, and anatomical keypoint extraction into a unified framework for automated cattle morphometric measurement and livestock monitoring.

1.4 Document Structure

This dissertation is organized into five chapters.

Chapter 1 introduces the research background, motivation, problem statement, research objectives, and the main contributions of the study. It also highlights the importance of computer vision techniques for automated cattle morphometric measurement and their potential applications in precision livestock farming.

Chapter 2 presents the literature review, providing an overview of existing research on laser-based calibration, region of interest (ROI) identification, animal segmentation, pose estimation, anatomical keypoint identification, and image-based cattle morphometric measurement. The chapter reviews the state of the art in these research areas and identifies the research gaps that motivated the development of the proposed framework.

Chapter 3 describes the proposed methodology, including the dataset preparation process, laser point detection and calibration, animal segmentation, anatomical keypoint identification, morphometric measurement estimation, and the evaluation methodology adopted in this study.

Chapter 4 presents the experimental results and performance evaluation of the proposed framework. The results are organized according to the three principal stages of the methodology, namely laser point detection, animal segmentation, and anatomical keypoint identification with morphometric measurement estimation. Both qualitative and quantitative evaluations are presented to assess the effectiveness, accuracy, and robustness of the proposed framework.

Finally, Chapter 5 summarizes the main findings of the research, discusses the contributions and limitations of the proposed framework, and presents recommendations and future research directions for automated cattle morphometric analysis.

Chapter 2

Literature Review

2.1 Region of Interest Identification

Region of Interest (ROI) refers to a specific area within an image that is considered relevant or significant for further analysis. Within computer vision applications, the importance of automatic ROI identification is crucial as it allows for the reduction of pixels that need to be processed. By focusing computational efforts only on the relevant regions of the images, there is an increase in processing speed while maintaining the same level of accuracy as processing the entire image. Furthermore, ROI identification helps reduce background noise and irrelevant visual information that may negatively affect subsequent image analysis tasks, thereby improving the robustness and reliability of the overall computer vision pipeline (Malek et al. 2025). ROI algorithms implement techniques such as threshold-based, feature-based, machine learning (ML), and deep learning methods to identify areas of interest.

In livestock monitoring applications, ROI identification plays an important role in localizing cattle within complex farm environments before subsequent analysis is performed. Accurate identification of the animal region reduces background interference and enables computer vision models to focus on the relevant anatomical structures of the bovine body. Andrew et al. (Andrew, Greatwood, and Burghardt 2017) proposed a deep learning-based framework for the visual localization of Holstein Friesian cattle using a Faster R-CNN architecture. The proposed system generated Regions of Interest (RoIs) in the form of bounding boxes that localized cattle from top-view and aerial images before further identification and analysis. Experimental results demonstrated a cattle localization accuracy of 99.3%, highlighting the effectiveness of deep learning-based ROI identification for livestock applications.

The identified ROI can subsequently be utilized for downstream tasks such as bovine segmentation, anatomical keypoint identification, morphometric measurement extraction, body condition assessment, and live weight estimation. Therefore, ROI identification serves as a fundamental pre-processing stage that improves both the efficiency and accuracy of computer vision-based cattle analysis systems.

2.2 Segmentation

Segmentation remains a challenging task in computer vision, as it involves addressing difficult issues such as feature representation, model design, and generalization across diverse real-world scenes (Y. Yu et al. 2023; T. Zhou et al. 2024; Hao, Zhou, and Guo 2020; Hongshan Yu et al. 2018). In contrast, image classification is concerned with recognizing the presence of an object class, whereas object detection additionally requires localization of the object within the image (Diwan, Anirudh, Tembhurne, et al. 2022). Deep learning models such as Region Based Convolutional Neural Network (**R-CNN**), Fast R-CNN, Faster R-CNN, and Mask R CNN are commonly employed to address this task. While these techniques are known for their efficiency, they may fall short in terms of precision for certain applications, as the resulting segmentations can exclude parts of the target object or mistakenly include background regions (Jessica Nye, L. M. Zingaretti, and Miguel Pérez-Enciso 2020). Due to high computational costs, the breakthrough of CNN applications happened not before it was possible to implement them on Graphics Processing Unit (**GPU**) to accelerate the training (Jennifer Salau and Krieter 2020). Bovine image segmentation has been widely investigated as a key component of automated livestock monitoring systems, where accurate delineation of the animal body is required for subsequent tasks such as pose estimation, behavior analysis, and body measurement. The complexity of farm environments and the variability in animal appearance have motivated the development of a wide range of segmentation approaches in the literature.

2.2.1 U-Net

U-Net is a convolutional neural network architecture designed for semantic image segmentation. It was introduced by Ronneberger et al. in 2015 (Kumar, Singh, and Sharma 2020). The model has been widely used because it can produce accurate pixel-level segmentation even with limited training data. The architecture follows an encoder–decoder structure in which the encoder extracts contextual features while the decoder reconstructs the segmentation mask (Ronneberger, Fischer, and Brox 2015). Skip connections between corresponding encoder and decoder layers allow the network to combine high-level semantic information with low-level spatial features, improving boundary localization and segmentation accuracy (Ronneberger, Fischer, and Brox 2015). One of the main strengths of U-Net is its ability to achieve accurate segmentation results even when the available training dataset is relatively small. This capability is largely attributed to its efficient architecture and the use of data augmentation strategies during training (Ronneberger, Fischer, and Brox 2015; Isensee et al. 2021). In addition, the relatively simple structure of the model makes it computationally efficient compared with deeper segmentation architectures, allowing it to be trained effectively using modern GPU-based deep learning frameworks (Joerg Salau and Krieter 2020). Despite its advantages, U-Net also presents several limitations. The architecture primarily relies on convolution operations with limited receptive fields, which may restrict its ability to capture long-range contextual information in complex scenes (Z. Zhou et al. 2020). As a result, segmentation performance may decrease when objects exhibit large scale variations or when

global context is required to distinguish between foreground and background regions. Furthermore, the increasing number of feature channels in both the encoder and decoder paths may lead to high memory consumption when processing high-resolution images (Isensee et al. 2021). These limitations have motivated the development of improved U-Net variants and other segmentation architectures that aim to enhance contextual feature representation and segmentation performance.

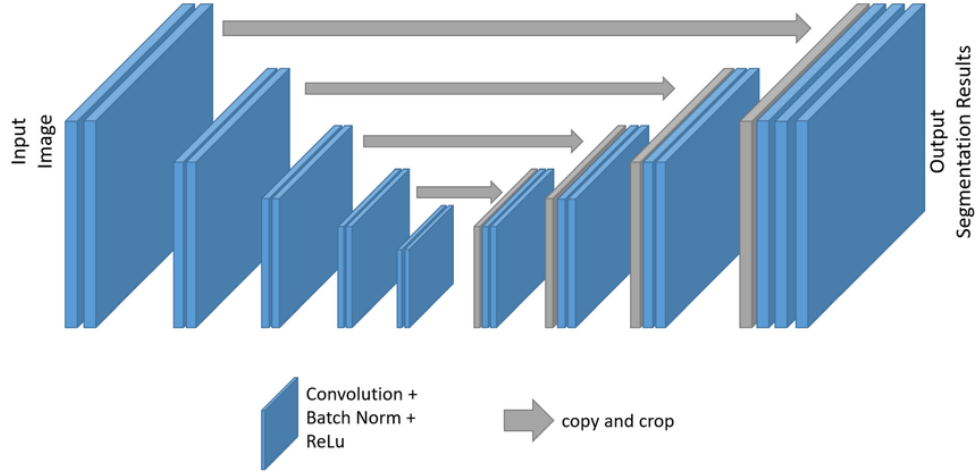


Figure 2.1: Architecture of U-Net for semantic segmentation. Extracted from: (Ronneberger, Fischer, and Brox 2015).

2.2.2 DeepLabV3+

DeepLabV3+ is a deep learning architecture designed for semantic image segmentation that improves the ability of convolutional neural networks to capture multi-scale contextual information. The model extends previous DeepLab architectures by introducing an encoder-decoder structure that combines atrous convolution with spatial pyramid pooling to enhance segmentation accuracy (Chen et al. 2018). In this architecture, the encoder extracts high-level semantic features using atrous convolution, which enlarges the receptive field without reducing the spatial resolution of feature maps, while the decoder refines the segmentation boundaries by incorporating low-level spatial features to produce detailed segmentation masks.

One of the main strengths of DeepLabV3+ is its capability to capture contextual information at multiple scales using the Atrous Spatial Pyramid Pooling (ASPP) module. The design allows the network to effectively handle objects with different sizes and complex backgrounds, which is particularly important in real world image segmentation tasks (Chen et al. 2018). In addition, the encoder-decoder framework improves the localization of object boundaries compared with earlier segmentation models. Several studies have applied DeepLabV3+ for livestock monitoring and agricultural applications. For instance, the author of (Feng et al. 2023) proposed an improved DeepLabV3+ model for cattle segmentation in complex farm environments by replacing the backbone network with MobileNetV2 and introducing additional feature fusion and attention mechanisms to enhance feature extraction capability.

Despite its strong performance, DeepLabV3+ also has several limitations. The use of atrous convolution and multi-scale feature extraction modules increases the computational complexity of the model, which may require significant computational resources for training and inference. Furthermore, segmentation performance may decrease when dealing with small objects or heavily occluded targets in complex scenes. These challenges have motivated further improvements to the DeepLab architecture and the development of alternative segmentation models that aim to enhance segmentation efficiency and robustness.

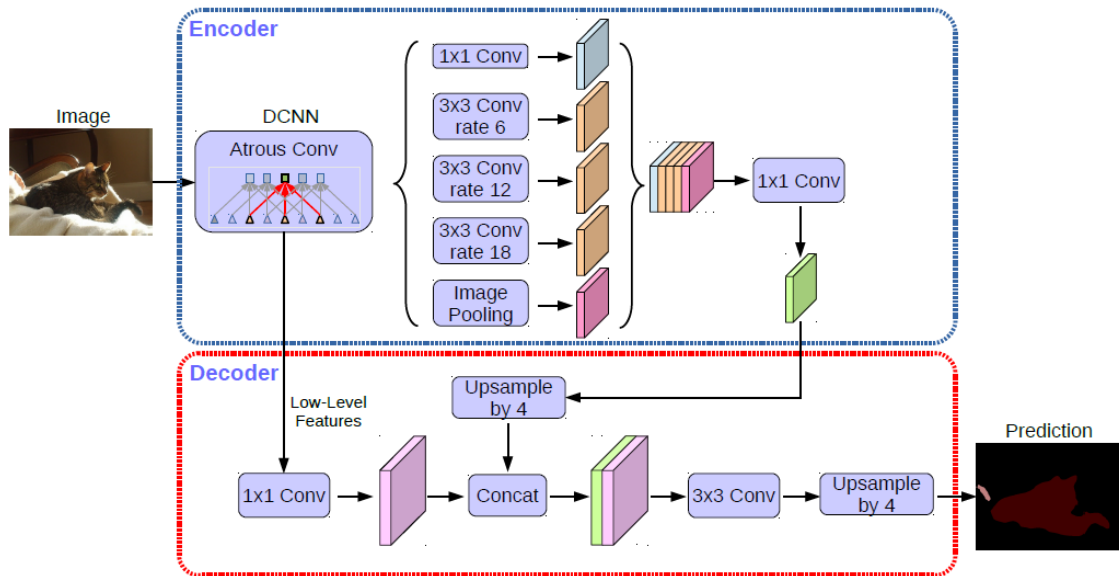


Figure 2.2: Encoder-Decoder Architecture of DeepLabV3+ Extracted from: (Feng et al. 2023)

2.2.3 Mask R-CNN

Mask R-CNN is a deep learning architecture developed for instance segmentation, which aims to detect objects in an image while simultaneously generating a segmentation mask for each detected instance. The method was introduced by He et al. (He et al. 2017a) as an extension of the Faster R-CNN object detection framework. The architecture follows a two-stage pipeline in which a Region Proposal Network (RPN) first generates candidate object regions, and a second stage performs object classification and bounding box regression for each proposed region. In addition to these tasks, Mask R-CNN introduces a parallel branch for predicting segmentation masks for each Region of Interest (RoI). The model also replaces the traditional RoIPool operation with the RoIAlign layer, which preserves exact spatial alignment between the feature maps and the input image. This improvement significantly enhances the quality of pixel-level segmentation by avoiding the quantization errors introduced by previous pooling operations (He et al. 2017a). One of the main strengths of Mask R-CNN is its ability to perform instance segmentation by identifying individual objects and generating precise masks for each object instance. The framework has demonstrated strong performance in complex scenes and has become a widely used baseline in many computer vision applications, including object detection, medical image analysis, and

autonomous systems (Kirillov et al. 2019). Despite its effectiveness, Mask R-CNN also presents several limitations. The architecture relies on a multi-stage detection pipeline, which increases computational complexity and requires substantial computational resources for both training and inference. Furthermore, the performance of the model may decrease when objects are heavily occluded or when detecting very small objects in cluttered environments (Minaee et al. 2021). These challenges have motivated the development of more efficient segmentation architectures and transformer-based segmentation models.

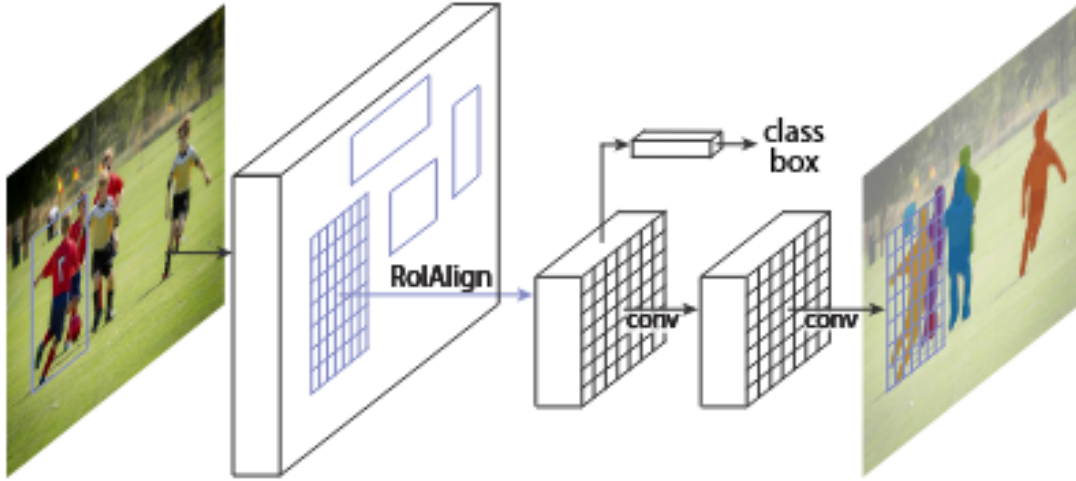


Figure 2.3: The Mask R-CNN framework for instance segmentation Extracted from: (He et al. 2017b).

2.2.4 Feature Pyramid Network (FPN)

Feature Pyramid Network (FPN), introduced by Lin et al. (Lin et al. 2017), is a widely adopted architecture for multi-scale feature representation in computer vision tasks such as image segmentation and object detection. FPN addresses the inherent limitations of conventional convolutional neural networks, where high-level feature maps provide strong semantic information but lack spatial resolution, while low-level feature maps retain spatial details but contain limited semantic context. To overcome this, FPN employs a bottom-up pathway, a top-down pathway, and lateral connections to effectively fuse features across multiple scales, generating semantically rich and high-resolution representations. The bottom-up pathway extracts hierarchical features from the encoder, while the top-down pathway, together with lateral connections, combines high-level semantic information with low-level spatial details, improving segmentation performance (Sun et al. 2023).

Following its introduction, numerous studies have focused on improving the original FPN structure. Path Aggregation Network (PANet) (S. Liu et al. 2018) extends FPN by introducing a bottom-up path augmentation to enhance information flow and improve localization capability. Similarly, Bi-directional Feature Pyramid Network (BiFPN) (Tan, Pang, and Le 2020) improves

feature fusion by incorporating bidirectional connections and learnable weights, enabling more efficient and scalable multi-scale feature integration.

Further advancements in FPN research have explored the integration of attention mechanisms to enhance feature representation. Variants such as Attention Aggregation Feature Pyramid Network (**A²-FPN**) and Scale-Sequence Feature Pyramid Network (**S²-FPN**) introduce attention-based modules to emphasize relevant spatial and channel features, improving segmentation accuracy, particularly in complex and multi-scale scenarios.

Recent research has extended the Feature Pyramid Network (FPN) architecture through attention-based modules and feature alignment techniques, improving multi-scale feature representation and semantic segmentation performance (Wang et al. 2024; Dong et al. 2026; H. Zhao et al. 2026). For example, SN-FPN introduces nested pyramid structures combined with self-attention mechanisms to enhance multi-scale feature consistency (Wang et al. 2024). Similarly, Adaptive FPN (AFPN) and MambaFPN integrate advanced context modeling techniques, including transformer-based and state-space representations, to improve long-range dependency modeling and computational efficiency.

Recent work by Dong et al. (Dong et al. 2026) further highlights the limitations of traditional FPN in handling semantic gaps between feature levels and the loss of local details. To address this, the authors propose AFPN, which enhances feature alignment and multi-scale fusion, achieving competitive performance on benchmark datasets such as Cityscapes, CamVid, and ADE20K.

Overall, these continuous improvements demonstrate that FPN remains a highly effective and evolving architecture. Its ability to efficiently fuse multi-scale features while maintaining computational efficiency makes it a strong state-of-the-art approach and a widely used benchmark for evaluating modern image segmentation methods.

2.2.5 SAM

The Segment Anything Model (SAM), developed by Meta AI Research, represents a significant breakthrough in computer vision, offering a robust framework for image and video segmentation. It is a versatile segmentation model designed to operate across a wide range of visual tasks. By leveraging a promptable architecture, SAM allows users to interactively segment objects using a variety of prompts, providing a level of flexibility and precision that surpasses traditional segmentation methods (Zhang et al. 2023). SAM builds upon transformer-based architectures, particularly Vision Transformers (ViT), which have demonstrated strong capability in capturing global contextual information (Dosovitskiy et al. 2021). A key contribution of SAM is the introduction of the SA-1B dataset, which contains approximately 11 million images and over 1 billion masks, making it significantly larger than existing segmentation datasets. The scale of this dataset allows SAM to achieve strong generalization performance and robustness across different domains. Similar large-scale dataset approaches have been shown to improve model performance in tasks such as object detection and segmentation (Lin et al. 2014a; Kuznetsova et al. 2020). Furthermore,

SAM demonstrates strong zero-shot transfer capabilities, performing competitively across multiple downstream tasks such as instance segmentation and object proposal generation without task-specific training. This aligns with the broader trend of foundation models in machine learning, where large-scale pretraining enables models to generalize effectively (Bommasani et al. 2021). Despite its advantages, SAM has limitations. It may struggle with fine boundary precision and small object segmentation, and its reliance on large-scale transformer architectures results in high computational requirements. These challenges are consistent with other transformer-based models in computer vision, which often trade efficiency for performance.

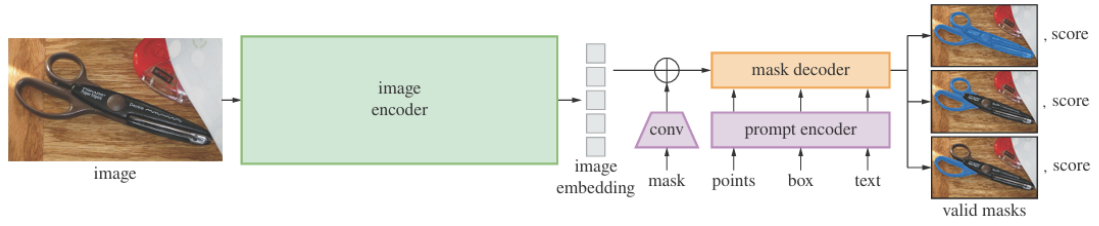


Figure 2.4: Architecture of the Segment Anything Model (SAM) showing the image encoder, prompt encoder, and mask decoder. Kirillov, Mintun, and Ravi (2023)

Table 2.1 shows that recent segmentation models achieve strong performance across a variety of datasets. U-Net remains popular because of its simple architecture and effectiveness on relatively small datasets. DeepLabV3+ improves contextual understanding through atrous spatial pyramid pooling, whereas Mask R-CNN provides instance-level segmentation. FPN enhances multi-scale feature extraction, and SAM demonstrates strong generalization capability through large-scale pretraining. However, these methods are designed primarily for generic segmentation tasks and do not explicitly incorporate laser-guided calibration or support anatomical keypoint extraction for cattle morphometric analysis. This limitation motivates the proposed framework.

Table 2.1: Comparison of Deep Learning-Based Segmentation Methods

Method	Ref.	Type	Dataset	Dataset Size	Metric	Value	Advantage	Disadvantage
U-Net	Ronneberger, Fischer, and Brox 2015	CNN (Enc-Dec)	ISBI, PhC-U373, DIC-HeLa	35; 30	Dice	0.80–0.90	Works well with limited data	Limited scalability
DeepLabV3+	Chen et al. 2018	CNN + ASPP	PASCAL VOC 2012, Cityscapes, MS COCO	1464/1449; 2975/500; 118000	Accuracy	93.98%	Strong multi-scale context	High computational cost
Mask R-CNN	He et al. 2017a	Instance Segmentation	MS COCO, PASCAL VOC	118000/5000; 10000	AP	0.35–0.40	Accurate object segmentation	Slow inference
FPN	Lin et al. 2017	Multi-scale CNN	MS COCO	118000	IoU	0.80	Good multi-scale feature extraction	High memory consumption
SAM	Kirillov et al. 2023	Transformer	SA-1B	11M images; 1B masks	IoU	90%+	Strong zero-shot generalization	Very high computational requirements

2.3 Laser-Based Reference Detection

Laser projection is widely used in computer vision to introduce a measurement reference into an image, as projected laser points typically appear as high-intensity, color-distinct regions that can be reliably detected even in complex scenes (Huelsmann, Buechner, and Van Gemmeren 2007). Vision-based systems exploit laser markers captured by calibrated cameras to infer object position, orientation, and geometric dimensions by relating the image coordinates of detected laser points to known geometric constraints (Sasaki, Sakai, et al. 2020).

Several studies have demonstrated that laser-based reference systems provide a reliable means of establishing the relationship between image coordinates and real-world dimensions. By exploiting the known physical distance between projected laser points, these systems enable accurate image scaling and support non-contact dimensional measurement in applications where direct physical measurement is impractical or undesirable (Pereira et al. 2019).

The visual appearance of a laser spot is generally characterized by a compact, bright region with sharp intensity gradients relative to the surrounding surface. However, accurate laser point localization remains challenging because the appearance of the projected laser is influenced by illumination conditions, surface reflectance, camera characteristics, and image noise. Consequently, numerous laser detection approaches have been proposed, including intensity thresholding, color-based segmentation, blob detection, contour analysis, and centroid estimation techniques (Zhang, Miao, and Yan 2025).

Accurate localization of laser reference points is fundamental to vision-based measurement systems because localization errors directly affect the estimation of geometric scale and reduce the accuracy of subsequent dimensional measurements. Therefore, considerable research has focused on improving the robustness of laser detection under varying environmental conditions, including changes in illumination, surface texture, shadows, reflections, and object motion (Zhang, Miao, and Yan 2025).

Despite the progress reported in the literature, reliable laser reference detection remains a challenging problem in real-world livestock environments. Variations in animal appearance, fur reflectance, environmental illumination, and animal movement can significantly affect the visibility and localization accuracy of projected laser points. These challenges continue to motivate research toward more robust laser detection techniques for vision-based livestock measurement applications.

2.4 Pose Estimation

Animal Pose Estimation (APE) is more difficult compared to human pose estimation, as people have the same key points distribution, appearance and movement pattern. However, the appearance, background environment, and size of different animals make them vary greatly, leading to a major challenge in estimating the pose (Jiang et al. 2022). The main goal of animal pose estimation is to identify a series of body keypoints which helps to understand the animal behavior,

part segmentation as well as for wildlife conservation. In fact, APE is useful in many practical daily applications ranging from farming to entertainment. Thus, many efforts have been made for animal pose estimation tasks due to community demand (Yang et al. 2022; Hang Yu et al. 2021a).

The effectiveness of human pose estimation has been shown in numerous studies and encouraging results were mainly attained because of the simplicity in the estimation. availability of large-scale, labeled human datasets in addition to the creation of sophisticated deep learning models. However, the lack of annotated animal pose data prevented these techniques from being used for animals. To address the scarcity of animal pose data, researchers have proposed cost-effective strategies, such as leveraging large-scale human pose datasets as prior knowledge (Cao et al. 2019). Although these approaches have achieved encouraging accuracy in animal pose estimation, the use of synthetic animal images as a substitute for real data remains challenging (Li and Lee 2021a; Shooter, Malleson, and Hilton 2021). Synthetic data generation is time-consuming and demands expert knowledge, while models trained on such data often fail to generalize effectively, particularly under self-occlusion and densely populated scenes.

Animal pose estimation has recently gained significant research interest, leveraging 2D and 3D models; however, the limited availability of large-scale annotated animal datasets, such as COCO (Lin et al. 2014b), poses a key challenge. Because humans and certain animal species, such as quadrupedal mammals, share similar skeletal structures, large-scale human pose datasets can be effectively leveraged to improve animal pose estimation by providing valuable prior knowledge. The study presented in (Cao et al. 2019) demonstrated that cross-domain adaptation can effectively improve performance in animal pose estimation tasks. They developed the Weakly- and Semi-Supervised Cross-Domain Adaptation (**WS-CDA**) strategy in order to transfer knowledge from human and animal data to unseen animals, and then, the Progressive Pseudo-Label-Based Optimization (**PPLBO**) is designed to boost model performance. The obtained results indicate that the accuracy of pose estimation at the human-level is achieved in animal estimation.

2.4.1 Unsupervised Domain Adaptation (Synthetic to Real data)

Due to the lack of animal labeled data, pseudo labels generated from synthetic data are widely used. The main challenge for these data is noise caused by domain shift. To the aim of reducing the domain gap between two different domains, i.e. synthetic and real data, the Multi-scale Domain Adaptation Module (**MDAM**) method is proposed by authors in (Li and Lee 2021a). In addition, a coarse-to-fine strategy was introduced to progressively generate new pseudo-labels to replace the initial ones. Accordingly, the model is trained in a two-stage manner: during the first stage, training relies on the initial pseudo-labels, and it gradually transitions to the newly generated pseudo-labels to mitigate overfitting. Experimental evaluations on the VisDA 2019 and TigDog datasets demonstrate that the proposed unsupervised domain adaptation approach significantly outperforms previously reported methods, while also exhibiting strong generalization capability to both unseen animal classes and unseen domains.

2.4.2 ScarceNet

The ScarceNet model was proposed by Chen Li and Gim Hee Lee (Li and Lee 2023a) to address animal pose estimation under scarce supervision. It exploits a limited amount of labeled data in conjunction with unlabeled samples to learn robust pose representations, demonstrating that competitive 2D animal pose estimation performance can be achieved even with minimal annotation. This approach highlights the potential of reducing annotation dependency in animal pose estimation tasks, particularly in scenarios where data collection is expensive or impractical.

Despite its effectiveness under limited supervision, ScarceNet remains focused on 2D pose estimation and still requires a certain number of annotated 2D keypoints during training. Moreover, it does not explicitly address the problem of 3D pose estimation or the inherent ambiguity associated with lifting monocular 2D poses to 3D space. As a result, while ScarceNet alleviates annotation scarcity, it does not resolve the challenges of learning 3D animal pose without paired 2D–3D supervision or real multi-view data, which are difficult to obtain in natural and wildlife settings (Silvia Zuffi et al. 2019; Rhodin, Salzmann, and Fua 2018).

In contrast to scarce-supervision methods such as ScarceNet (li2023scarcene), the method proposed in this thesis estimates 3D animal pose from monocular 2D observations without requiring paired 2D–3D annotations or real multi-view supervision, thereby overcoming the limitations of existing approaches and improving geometric plausibility in 3D pose estimation.

2.5 Anatomical keypoint identification

Early research on animal keypoint detection primarily targeted simple scenarios, typically involving animals captured in clean backgrounds with minimal environmental complexity (Y. Wang and L. Zhao 2026).

2.5.1 Early Vision-Based Methods

Early vision-based approaches for anatomical keypoint detection in livestock primarily relied on classical computer vision techniques and hand-crafted features. Early studies in pigs (Kashiha et al. 2014) used predefined body landmarks to monitor posture and movement, providing one of the first automated benchmarks for livestock behavior analysis. In dairy cows, top-view and side-view video frames were analyzed using edge detection, background subtraction, and contour-based feature extraction to detect standing, lying, and walking postures (Smith and Brown 2020). Similarly, the estimation of sheep posture and orientation leveraged thresholding, ROI tracking, and shape analysis to locate key anatomical landmarks (Jones and Patel 2019). These early methods enabled basic behavioral monitoring and facilitated preliminary welfare assessments, but they suffered from significant limitations. Sensitivity to illumination variations, shadows, and occlusions caused by other animals or farm infrastructure reduced detection reliability. Manual parameter tuning for camera angles, lighting conditions, and thresholds was often required to achieve acceptable results (Liu and Wang 2018; Oliveira and Silva 2017). Importantly, the accuracy of these methods was

limited, often failing to precisely localize anatomical keypoints under real-world conditions or complex animal poses. Reported errors in landmark detection could range from several centimeters to tens of centimeters, depending on the body region and environmental conditions, which is insufficient for precise behavioral or morphometric analyses (Kashiha et al. 2014; Smith and Brown 2020).

Despite their limitations, early vision-based methods established the foundation for subsequent deep learning approaches by demonstrating the importance of anatomical landmarks and structured keypoint detection for livestock monitoring and analysis (Kashiha et al. 2014; Smith and Brown 2020). They also provided valuable datasets and evaluation benchmarks that have been widely adopted and extended in later automated keypoint detection studies.

2.5.2 Deep Learning-Based Approaches

The emergence of deep learning has transformed anatomical keypoint detection in livestock, providing significant improvements in accuracy, robustness, and scalability over classical vision-based approaches. Among the first frameworks to leverage these advances, transfer learning-based methods such as *DeepLabCut* have been applied to pigs, cattle, and sheep, allowing automatic identification of key anatomical landmarks with minimal labeled data (Mathis et al. 2018; Y. Liu et al. 2025). By utilizing pretrained convolutional neural networks, these approaches can extract rich feature representations, reducing dependency on large annotated datasets and enabling practical deployment in farm environments.

Beyond simple landmark detection, convolutional neural network (CNN)-based models have been specifically employed for lameness assessment in dairy cows, targeting spine and limb landmarks to support objective behavioral and welfare monitoring (Hansen et al. 2018). Multi-animal pose estimation frameworks have further extended these approaches, allowing simultaneous tracking of multiple individuals within herds. These methods improve temporal consistency and facilitate analyses of social interactions, activity patterns, and locomotion dynamics (Perneel and Smith 2025; Parivendan and Rao 2025). Temporal and trajectory-based models that incorporate spatio-temporal keypoint information have proven particularly effective for capturing subtle behavioral patterns and complex locomotor interactions (Wei and Zhang 2023).

Deep learning approaches have also been adapted for other livestock species and specialized tasks. For instance, sheep pose estimation integrates CNNs with object detection to extract precise body measurements from anatomical keypoints (Bati and Khan 2025), while transformer-based architectures have been applied to yak heifers to achieve robust performance in complex environmental conditions (Yakheifer and Singh 2025). In addition, face-specific keypoint detection for cows, using architectures such as YOLOv7-Pose, has enabled automated monitoring of head pose and behavior, providing valuable tools for welfare assessment and feeding behavior analysis (Huang and Lee 2024). Ensemble learning techniques that combine multiple motion features extracted from keypoints have further enhanced lameness detection accuracy across variable farm conditions (Smith and Patel 2025).

Despite these successes, several challenges remain. Deep learning-based methods are often sensitive to multi-animal occlusion, varying lighting conditions, differences in breed and body size, and complex farm environments, which can limit generalization (Y. Wang and L. Zhao 2026). Moreover, most frameworks require dataset-specific fine-tuning, reducing their plug-and-play usability across different farms. Current research efforts are therefore focused on increasing model robustness, integrating temporal and 3D information, and exploring semi-supervised or self-supervised learning strategies to reduce annotation dependency while maintaining high accuracy (Mathis et al. 2018; Y. Liu et al. 2025).

Overall, deep learning approaches represent a significant advancement over classical methods, offering improved precision, adaptability, and scalability in livestock keypoint detection, but careful attention to environmental factors, dataset variability, and model generalization remains critical for real-world applications.

2.5.3 CNN-Based Keypoint Detection

Convolutional Neural Network (CNN) have become a cornerstone for anatomical keypoint detection in livestock, particularly for tasks requiring precise localization of skeletal and limb landmarks. In dairy cattle, CNN-based models have been extensively applied to detect spine and limb keypoints to support lameness assessment (Hansen et al. 2018). By learning hierarchical spatial features from image data, these networks can automatically identify subtle postural deviations that indicate potential health or welfare issues. Studies have demonstrated that CNN models can achieve high accuracy in controlled farm environments, reliably localizing anatomical points under standardized lighting and camera setups. Beyond lameness detection, CNN-based keypoint detection has been extended to monitor locomotion, behavior, and social interactions. Multi-animal tracking frameworks based on CNNs enable simultaneous analysis of group-housed livestock, facilitating studies on herd movement patterns and activity dynamics (Perneel and Smith 2025; Parivendan and Rao 2025). Temporal modeling, when combined with CNN feature extraction, further allows researchers to capture dynamic behaviors over time, such as walking gait cycles, lying-to-standing transitions, and feeding postures (Wei and Zhang 2023). Despite these advances, CNN-based approaches face several challenges in practical farm settings. Performance can degrade due to occlusion, overlapping animals, or variable viewpoints that frequently occur in barns (Y. Wang and L. Zhao 2026). Changes in lighting, shadows, and environmental clutter also reduce detection robustness. Additionally, most CNN models require careful dataset-specific fine-tuning, limiting their generalization across different breeds, barn layouts, or species (Y. Liu et al. 2025). These limitations have motivated the integration of advanced architectures, including multi-scale CNNs, attention mechanisms, and temporal-spatial modeling, to improve both precision and robustness under realistic farm conditions.

In summary, CNN-based keypoint detection has proven highly effective for automated anatomical landmark localization in livestock, enabling applications ranging from lameness assessment to

behavioral monitoring. However, its practical deployment requires addressing environmental variability, occlusion, and generalization challenges, which are critical considerations for large-scale farm implementations.

2.5.4 Face and Behavior-Specific Keypoint Detection

Face and behavior-specific keypoint detection provides critical insights into livestock welfare, stress, feeding, and social interactions. In dairy cows, YOLOv7-Pose architectures have been used for facial landmark detection and head pose recognition, enabling automated monitoring of feeding, lying, and social behaviors (Huang and Lee 2024). Ensemble learning methods that integrate multiple motion descriptors from limbs, spine, and facial keypoints have improved lameness detection accuracy and robustness (Smith and Patel 2025).

These techniques have also been applied to other livestock species. In pigs, facial and body motion analysis is used to detect feeding patterns and aggression, while in sheep, posture and locomotion are monitored through keypoint-based motion descriptors (Y. Liu et al. 2025; Bati and Khan 2025). Limitations include multi-animal occlusion, lighting variation, barn layout differences, and computational complexity. Nevertheless, these approaches represent a significant advancement over classical methods, providing precise, automated, and welfare-relevant monitoring tools across multiple livestock species.

Figure 2.5 provides a summary of the main anatomical keypoint identification approaches in livestock, highlighting the progression from traditional vision-based methods to modern deep learning and behavior-specific keypoint detection techniques.

Anatomical Keypoint Identification in Livestock

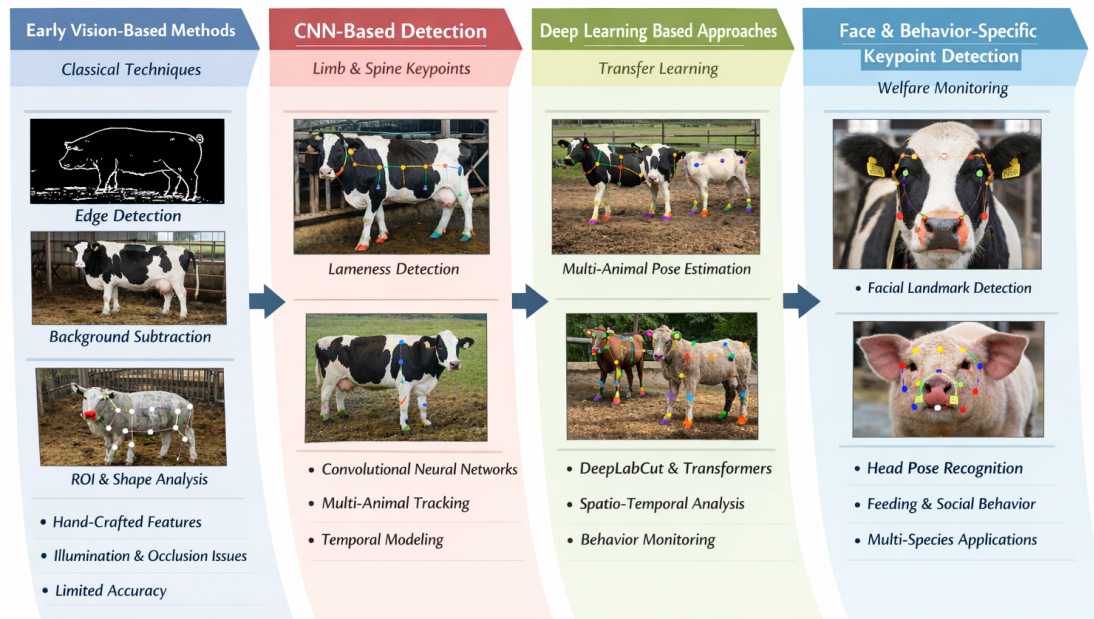


Figure 2.5: Overview of anatomical keypoint identification approaches in livestock, including classical vision-based methods, CNN-based detection, deep learning approaches, and behavior-specific keypoint analysis

2.6 Research Gap

The literature demonstrates substantial progress in the application of computer vision and deep learning techniques for livestock monitoring, including animal localization, image segmentation, pose estimation, anatomical keypoint identification, and morphometric analysis (Andrew, Greatwood, and Burghardt 2017; Mathis et al. 2018; Li and Lee 2023a). Recent advances in deep learning architectures such as U-Net, DeepLabV3+, Mask R-CNN, Feature Pyramid Networks (FPN), and the Segment Anything Model (SAM) have significantly improved segmentation accuracy and robustness across a wide range of computer vision applications (Ronneberger, Fischer, and Brox 2015; Chen et al. 2018; He et al. 2017b; Lin et al. 2017; Kirillov et al. 2023).

Despite these advancements, several limitations remain. Many existing studies investigate animal segmentation, pose estimation, anatomical landmark detection, and body measurement extraction as separate tasks rather than integrating them into a unified computer vision framework. Consequently, the interaction between these components and their combined impact on measurement accuracy are often overlooked. Furthermore, while considerable attention has been devoted to segmentation and animal pose estimation, comparatively fewer studies have explored anatomical keypoint identification for automated morphometric measurement extraction and livestock body dimension estimation (Mathis et al. 2018; Li and Lee 2023a).

Another important limitation is that several existing approaches rely on specialized sensing equipment, depth cameras, multi-camera systems, or controlled imaging environments, thereby increasing system complexity and limiting scalability in practical farm deployments (A. Ruchay et al. 2021; J. Nye, P. Zingaretti, and M. Pérez-Enciso 2020). In addition, segmentation and pose estimation models often experience performance degradation under challenging farm conditions characterized by varying illumination, animal pose variation, occlusions, and complex backgrounds.

Moreover, the integration of laser-based calibration with anatomical keypoint identification and image-based morphometric measurement extraction remains relatively underexplored in livestock computer vision research. Existing studies typically depend on manual reference measurements or dedicated hardware systems, reducing flexibility and increasing operational costs.

Therefore, there remains a need for an integrated computer vision framework capable of combining bovine segmentation, laser-based reference detection, anatomical keypoint identification, contour extraction, and morphometric measurement estimation using RGB images acquired under practical farm conditions. Addressing these challenges can contribute to the development of scalable, non-contact, and cost-effective livestock monitoring systems for precision livestock farming.

The present research addresses these limitations by proposing a unified framework that integrates cattle segmentation, laser-guided calibration, contour extraction, anatomical keypoint identification, and morphometric measurement estimation within a single processing pipeline. Unlike previous studies that primarily focus on individual tasks, the proposed approach combines multiple computer vision components to enable automated livestock body measurement under real-world farm conditions.

2.7 Summary

Table 2.2 summarizes the main categories of livestock image segmentation approaches, highlighting their reported performance, limitations, and applicability to complex farm environments.

Table 2.2: Comparison of livestock image segmentation methods: results, limitations, and representative studies

Method Category	Reported Results	Limitations / Gaps	Representative Studies
Classical Image Processing	Moderate accuracy in controlled settings; strong degradation under illumination changes, shadows, and cluttered backgrounds	Highly sensitive to lighting and background variation; poor generalization to real farm environments	(Porto and Smith 2013; Kashiha, Bahr, and Wiegand 2014)
Model-/Shape-Based Methods	More consistent silhouettes in constrained scenes	Limited robustness to posture changes, occlusion, and overlapping animals	(Tangorra and Rossi 2013)
Semantic Segmentation (CNN-based)	High pixel-level accuracy (IoU often >80%) in single-animal datasets	Large annotation requirement; reduced performance under severe occlusion or multi-animal interaction	(Wei and Zhao 2023; Yu and Li 2023)
Instance Segmentation	Accurate individual masks in multi-animal scenes; supports downstream pose estimation	Computationally demanding; sensitive to training data diversity	(Nye and Turner 2020; Sasaki and Tanaka 2025)
Panoptic Segmentation	Unified semantic and instance-level scene understanding	Limited livestock-specific validation; high model and annotation complexity	(Kirillov, He, and Girshick 2019)
Multi-view / 3D Segmentation	Improved accuracy under occlusion and animal overlap	High system complexity; multi-camera setup and calibration required	(Psota and Kriška 2020)

As shown in Table 2.3 most existing animal pose estimation approaches rely on large annotated datasets, synthetic supervision, or multi-view camera configurations. Such dependencies restrict their applicability in commercial farming environments, where collecting extensive annotations and installing multiple cameras can be costly and labor-intensive.

Table 2.3: Comparison of animal pose estimation approaches: performance and limitations

Approach	Performance Trends	Limitations / Gaps	Representative Studies
Cross-domain Adaptation	Transfers knowledge from large-scale human pose datasets to improve animal pose estimation accuracy	Limited availability of animal labels; domain gap between humans and animals; reduced performance for species dissimilar to the training set	(Cao2019)
Synthetic-to-Real Domain Adaptation	Improves generalization by adapting models trained on synthetic data to real animal images	Pseudo-label noise; persistent domain shift between synthetic and real data	(Li and Lee 2021b)
Fully Supervised Deep Learning	High accuracy on large benchmarks such as AP-10K	Requires extensive labeled datasets; limited generalization to rare poses and unseen species	(Hang Yu et al. 2021b, 2021c)
Scarce Supervision (ScarceNet)	Competitive pose estimation accuracy achieved with minimal labeled data	Residual pseudo-label noise; unreliable samples may limit final performance	(Li and Lee 2023b)
Multi-view 3D Pose Estimation	Accurate 3D pose reconstruction in controlled environments	Requires multiple calibrated cameras; impractical for large-scale farm or wildlife deployment	(Sergio Zuffi et al. 2019; Rhodin et al. 2018)
Monocular 3D Pose Estimation	Improves geometric plausibility without paired 2D–3D supervision	Single-view depth ambiguity; may underperform compared to multi-view approaches; limited cross-species generalization	(Rhodin et al. 2018; Sergio Zuffi et al. 2019; X. Wang and Y. Zhao 2025)

Table 2.4 shows that current anatomical keypoint detection methods achieve high accuracy in controlled settings but often rely on large annotated datasets and struggle with occlusions, varying viewpoints, and complex farm environments. These limitations highlight the need for approaches that are both precise and practical to deploy in real-world livestock conditions.

Table 2.4: Comparison of anatomical key point identification approaches in livestock vision systems

Method Category	Reported Results (Metrics)	Limitations / Gaps	Representative Studies
Early Vision-Based Methods	Moderate landmark localization accuracy in controlled environments; evaluated using localization error and detection rate, typically achieving below 70% accuracy.	Highly sensitive to illumination changes, occlusion, and background clutter; limited robustness and scalability in real farm environments.	(Cangar et al. 2008; Viazzi et al. 2014)
Deep Learning-Based Approaches	Significant improvement in anatomical landmark detection; commonly evaluated using Percentage of Correct Keypoints (PCK) and mean Euclidean error, with reported PCK values exceeding 85–90%.	Requires large annotated datasets; high computational cost during training; limited cross-species generalization.	(Mathis et al. 2018; Peng, Liu, and Zhang 2024)
CNN-Based Keypoint Detection	High-precision landmark detection using CNN backbones; evaluated using PCK@0.05–0.1, mean Average Precision (mAP), and RMSE, achieving PCK above 90% in single-animal scenarios.	Performance degrades with occlusion and overlapping animals; real-time deployment on edge devices remains challenging.	(Barbal, Bahr, and Berckmans 2023; Peng, Liu, and Zhang 2024)
Face and Behavior-Specific Keypoint Detection	Accurate detection of facial and behavior-related landmarks; evaluated using Normalized Mean Error (NME), PCK, and classification accuracy, often exceeding 90% for frontal facial landmarks.	Limited robustness to side views and occlusions; behavior annotation is labor-intensive; reduced performance in group housing conditions.	(Hansen, Smith, and Smith 2021; Zhang, Li, and Wang 2023)

Chapter 3

Methodology

3.1 Data Preparation

This section describes the cattle image dataset used in this research, as well as the preprocessing and annotation procedures applied to prepare the data for model development and evaluation.

The dataset used in this study was provided within the scope of the project *Medida 1. “Inovação” do PDR 2020 — Programa de Desenvolvimento Rural do Continente — Grupo Operacional Preservar a Qualidade na Carne Arouquesa (PDR2020–101–031094)* (PDR2020 2020). The project aims to promote innovation and technological solutions for the preservation and assessment of Arouquesa cattle production systems and meat quality. The dataset was made available for research purposes and serves as the basis for the development and evaluation of the computer vision methods proposed in this dissertation. This section describes the cattle image dataset provided for this research, as well as the preprocessing and annotation procedures applied to prepare the data for model development and evaluation. The dataset was utilized for animal segmentation, laser point detection, pose estimation, and morphometric analysis under realistic livestock farming conditions.

3.1.1 Image Acquisition Setup

Image recordings were obtained at the Portugal Livestock Research Centre’s dairy barn. Side-view RGB images of individual cattle were captured using a consumer-grade smartphone equipped with a built-in camera. The smartphone was mounted on a rigid support bracket together with two red laser pointers fixed at a known horizontal distance of 32 cm. These laser pointers projected visible laser dots onto the animal body surface, providing spatial reference markers for pixel-to-physical measurement calibration.

Image acquisition was carried out in real livestock environments, primarily within indoor cattle barns. These environments provided indirect natural illumination, ensuring consistent visibility of both the animal body and the projected laser points. During image capture, the camera was maintained approximately parallel to the animal body plane to preserve a consistent lateral perspective.

Depending on animal positioning and accessibility within the pen, each subject was photographed from either the left or right side. To enhance dataset diversity and robustness, images were acquired under varying environmental conditions, including different wall materials, floor surfaces, and lighting configurations. This variability reflects realistic farm conditions and supports the generalization of the proposed methods.

3.1.2 Data Annotation and Ground Truth

Each image in the dataset was manually annotated to establish reliable ground truth data for laser point detection. The annotations correspond to one or two visible laser points projected onto the animal body using the dual-laser system. The physical distance between the laser points was fixed at 32 cm and served as a reference for scale estimation.

Manual annotation was performed by recording the pixel coordinates of the visible laser points. For each laser point, the ground truth coordinates were stored in the form (X, Y) , where X denotes the horizontal pixel location and Y denotes the vertical pixel location relative to the top-left corner of the image. When both laser points were visible, the coordinates (X_1, Y_1) and (X_2, Y_2) were recorded. In cases where only one laser point was visible due to occlusion or camera angle, only a single coordinate pair was annotated.

In addition to laser point coordinates, metadata including the image filename, animal identification, acquisition date, acquisition time, and corresponding body weight measurements were recorded for each sample. These annotations serve as ground truth references for evaluating the accuracy of the automatic laser point detection algorithm proposed in this research. An example of an annotated input image from the dataset is shown in Figure 3.1.



Figure 3.1: Cattle images with different backgrounds from the used dataset

3.1.3 Preprocessing

The dataset consists of 460 RGB images of young cattle acquired in an indoor livestock housing environment. All images were captured from a lateral (side-view) perspective at a fixed spatial resolution of 4608×3072 pixels. The high image resolution preserves fine anatomical details and projected laser-point information, supporting accurate cattle segmentation, landmark localization, and morphometric measurement extraction. During image acquisition, two laser points were projected onto each animal to provide a reference for subsequent image-based measurements. The data were collected under controlled conditions to ensure consistency in image quality, animal characteristics, and camera viewpoint while maintaining sufficient background variability for robust model development and evaluation.

The main characteristics of the dataset are summarized as follows:

- Number of images: 460
- Image resolution: 4608×3072 pixels
- Color space: RGB
- Viewpoint: Lateral (side view)
- Environment: Indoor livestock housing

- Subjects: Young cattle of similar breed and coloration
- Laser projection: Two red laser points per image

3.1.4 Morphological Measurements Dataset

In addition to the image dataset, a morphometric dataset containing manually acquired anatomical measurements was used as ground-truth in this study. The measurements were collected during cattle weighing sessions and comprise 66 animals. The dataset includes animal identification, sex, age, live weight, and standard morphometric measurements recorded in centimeters. These manually collected measurements provide the reference values for developing and validating the proposed image-based morphometric measurement framework.

The morphometric dataset contains ten standard anatomical measurements that describe different aspects of cattle body morphology. As illustrated in Figure 3.2, these measurements correspond to standard anatomical traits commonly adopted for livestock characterization and body conformation assessment. Their definitions and anatomical descriptions are summarized in Table 3.1.

Although the proposed framework is designed to estimate all anatomical measurements contained in the dataset, this dissertation focuses on five representative morphometric traits: withers height (a_1), height at tail insertion (ANC), chest depth (a_4), body length (b_1), and rump length (b_2). These measurements were selected because they represent the principal vertical and longitudinal dimensions of the animal and are widely used for livestock growth assessment, body conformation analysis, and morphometric characterization. The remaining measurements belong to the original morphometric protocol and can be estimated using the same methodology, but they are not included in the implementation and experimental evaluation presented in this dissertation.

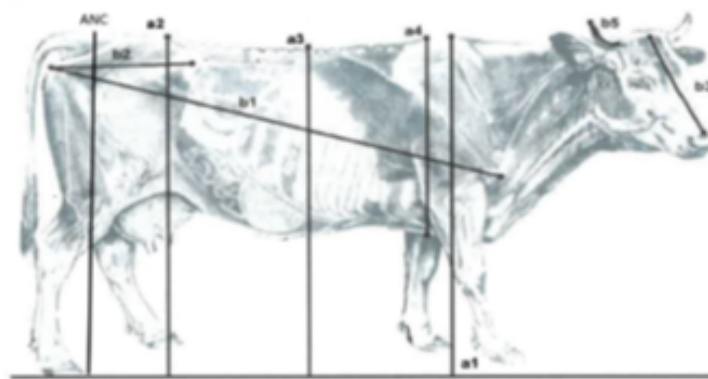


Figure 3.2: Standard anatomical measurements collected during dataset acquisition. The proposed framework estimates five representative morphometric traits: withers height (a_1), height at tail insertion (ANC), chest depth (a_4), body length (b_1), and rump length (b_2). The remaining measurements belong to the original morphometric protocol and are not considered in this study.

Table 3.1 defines each anatomical measurement included in the morphometric dataset. These definitions establish the reference used throughout this dissertation for anatomical landmark localization and morphometric measurement estimation.

Table 3.1: Definition of the anatomical measurements contained in the morphometric dataset.

Symbol	Measurement	Definition
a_1	Withers Height	Vertical distance from the highest point of the withers to the ground.
ANC	Height at Tail Insertion	Vertical distance from the ground to the base of the tail immediately after the sacral vertebrae.
a_4	Chest Depth	Vertical distance from the withers to the lowest point of the chest.
b_1	Body Length	Straight-line distance between the scapulo-humeral joint (shoulder) and the ischial tuberosity.
b_2	Rump Length	Straight-line distance between the tip of the ilium and the ischial tuberosity.
c_1	Trunk Width	Straight-line width of the trunk measured behind the shoulders.
c_2	Bi-Iliac Width	Straight-line distance between the outermost points of the left and right ilium bones.
c_3	Bi-Ischial Width	Straight-line distance between the outermost points of the left and right ischial tuberosities.
LCF	Coxo-Femoral Width	Distance between the left and right coxo-femoral (hip) joints.
d_1	Thoracic Perimeter	Thoracic circumference measured immediately behind the shoulders.

A representative sample of the morphometric dataset is presented in Table 3.2. The table includes animal characteristics together with the manually acquired anatomical measurements that served as the ground-truth reference for the development and quantitative validation of the proposed framework.

Table 3.2: Sample of the ground-truth dataset containing animal characteristics and morphometric measurements used in this study.

Animal	Sex	Age (days)	Weight (kg)	a_1	ANC	a_4	b_1	b_2	c_1	c_2	c_3	LCF	d_1
7	M	100	81.0	75.0	78.5	37.5	94.0	28.0	19.0	20.0	14.0	20.0	100.5
23	M	203	215.0	112.5	115.0	52.0	125.0	38.0	29.0	33.0	23.0	31.0	136.0
34	M	219	243.5	108.0	115.5	51.0	128.0	38.0	32.0	34.0	22.0	30.0	146.0
45	M	218	213.0	108.0	115.0	50.0	128.0	39.0	29.0	31.5	22.0	29.5	139.0
52	M	125	161.5	101.5	109.5	45.0	112.0	35.0	25.0	30.0	20.0	27.0	126.0
64	M	97	56.5	73.0	77.0	34.0	85.0	25.0	16.5	17.0	12.0	17.5	89.0
66	M	141	180.5	107.5	113.5	46.5	120.5	38.0	26.5	32.0	21.0	27.0	133.0
70	M	159	199.0	100.0	105.0	48.0	121.0	33.0	28.0	31.0	15.0	28.0	133.0
71	M	174	128.0	88.0	94.0	46.0	114.0	34.0	21.0	24.0	14.0	24.0	117.0
73	M	107	82.0	75.0	79.0	37.0	96.0	28.0	20.5	22.0	16.0	20.0	101.0
79	M	143	116.5	84.0	91.0	41.0	96.0	31.0	23.0	22.0	14.0	21.5	112.0
81	M	216	230.0	110.0	115.5	51.0	133.0	42.0	31.0	34.0	21.0	29.0	143.0
84	M	175	125.0	85.0	94.0	41.5	110.0	34.0	24.0	24.0	18.0	23.0	117.0
86	M	162	171.0	100.0	108.0	49.0	120.0	35.0	25.0	28.0	16.0	25.0	126.0
88	M	155	157.0	90.0	105.0	43.0	116.0	37.0	24.5	26.5	19.0	25.0	120.5

3.2 Proposed Framework

The objective of this dissertation is to develop a computer vision framework capable of automatically extracting morphometric measurements from monocular RGB images of cattle acquired under real livestock farming conditions. The proposed framework aims to provide a non-contact, low-cost, and scalable alternative to conventional manual measurement procedures by integrating image segmentation, laser-based calibration, anatomical landmark localization, and morphometric measurement extraction within a unified processing pipeline.

The framework receives a side-view RGB image of a cattle specimen as input and produces morphometric measurements expressed in real-world units as output. The proposed approach is designed to operate under practical farm conditions where background clutter, illumination variability, animal pose variation, and environmental complexity may affect image quality.

The proposed framework consists of three main stages:

1. Animal body segmentation;
2. Laser point detection and scale estimation;
3. Anatomical keypoint identification and distance calculation.

Initially, the animal body is isolated from the surrounding background through an image segmentation process. This stage generates a binary representation of the cattle silhouette that serves as the basis for subsequent anatomical analysis.

Simultaneously, the projected laser markers are identified and used to estimate a scale factor relating image coordinates to real-world dimensions. The known physical distance between the

laser projections provides a geometric reference that enables the conversion of image measurements into metric units.

Following segmentation and scale estimation, anatomical keypoints corresponding to relevant body structures are identified. These landmarks provide a structured representation of cattle morphology and establish the geometric relationships required for morphometric assessment.

Finally, morphometric measurements are estimated using the identified anatomical landmarks and converted from pixel units to real-world dimensions using the laser-based calibration factor.

The overall processing pipeline adopted in this dissertation is illustrated in Figure 3.3. The proposed framework integrates all computational stages into a unified workflow for automated cattle morphometric analysis.

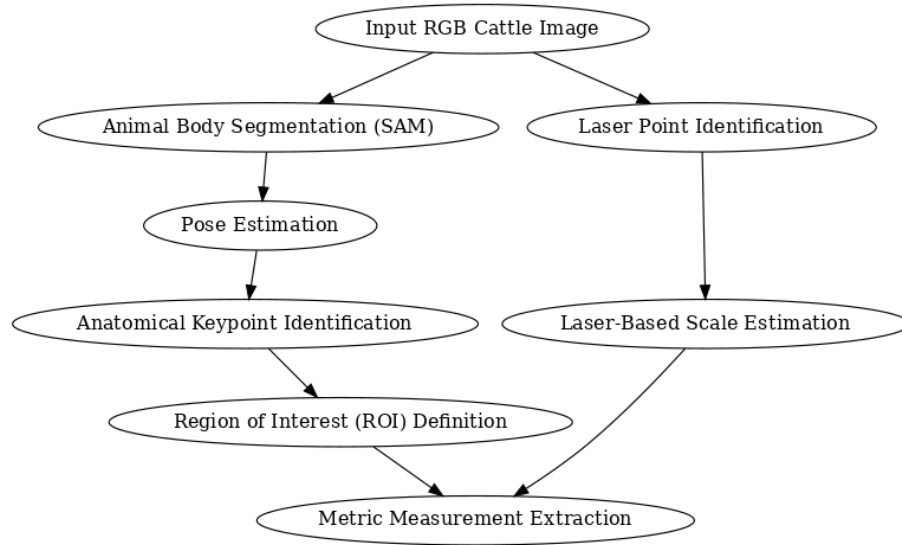


Figure 3.3: Overview of the proposed framework for cattle morphometric analysis.

The expected outputs of the proposed framework include segmented animal masks, detected laser-point locations, anatomical keypoint coordinates, and morphometric measurements corresponding to withers height (a_1), waist height (ANC), chest depth (a_4), body length (b_1), and rump length (b_2).

Figure 3.4 presents the overall organization of the proposed framework. The methodology is structured into three main stages: laser point identification, animal segmentation, and anatomic keypoint identification with distance calculation. These stages operate sequentially to enable the extraction of morphometric measurements from RGB cattle images. The laser point identification stage provides the scale reference required for metric estimation, the animal segmentation stage isolates the cattle body from the background, and the final stage identifies anatomical landmarks and computes the target morphometric measurements.

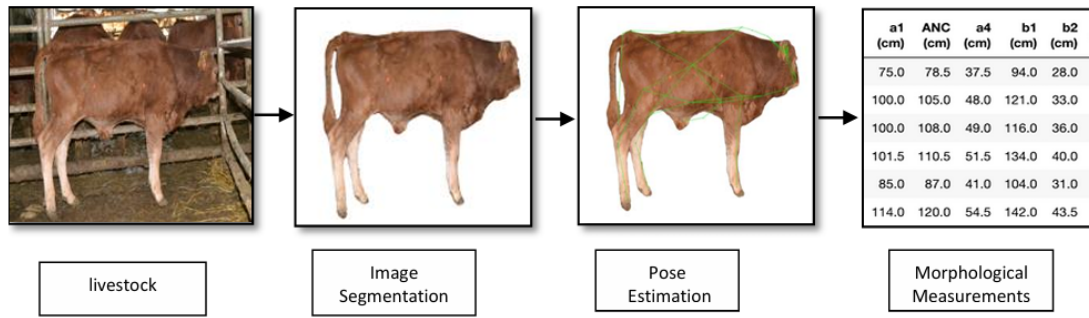


Figure 3.4: Overall framework adopted in this dissertation. The methodology consists of three main stages: (1) laser point identification, (2) animal segmentation, and (3) anatomic keypoint identification and distance calculation. The detected laser points provide the scale reference, the segmentation stage isolates the animal body, and the identified anatomical landmarks are used to estimate morphometric measurements.

3.3 Framework Implementation

3.3.1 Laser Point Identification

Laser point identification is conducted using a classical image-processing algorithm designed to detect projected laser points on the animal body surface. The method exploits the distinctive colour characteristics of the laser projections to isolate laser pixels from the surrounding scene and subsequently estimate a metric reference for physical measurement calibration.

The projected laser points appear as bright red regions in the acquired images, as illustrated in Figure 3.5. However, variations in illumination conditions, laser intensity, surface reflectance, and camera acquisition settings can significantly affect their RGB colour representation. To improve robustness against these variations, the input RGB image is transformed into the perceptually uniform CIE LAB colour space, which provides a more reliable representation of perceptual colour differences and enables consistent comparison between image pixels and predefined laser colour references.

Following colour-space transformation, colour-based thresholding is applied to extract candidate laser regions according to their similarity to the reference laser colours. The resulting binary mask is subsequently refined through morphological filtering to suppress noise, remove spurious detections, and improve the continuity of valid laser regions. Connected component analysis is then performed to identify individual laser regions, from which centroid coordinates and bounding-box information are computed. These detections provide the spatial reference required for scale estimation and the conversion of image-based measurements into real-world dimensions.



Figure 3.5: Reference laser colour samples extracted from the dataset..

1) *Reference Color Encoding:* To characterize the appearance of the projected laser points, fifteen representative RGB colour samples were manually extracted from laser-point regions in the acquired images. These samples were selected to capture the variability of laser appearance under different illumination and imaging conditions and were subsequently used as reference colours for laser-pixel classification. Each reference colour $RGB = (R, G, B)$ is transformed into the CIE LAB colour space using a standard colour-space conversion function:

$$LAB = RGB_to_LAB(R, G, B) \quad (3.1)$$

A. Image Segmentation via CIEDE2000 Thresholding

Segmentation of laser points within the captured images is an important step in the proposed detection pipeline. To achieve robust segmentation, the method relies on perceptual color similarity in the CIE LAB color space, which offers a more uniform representation of human color perception compared to the traditional RGB model. This allows for more accurate differentiation between laser dots and surrounding features, even under varying lighting conditions or surface textures. The segmentation process begins by transforming both the input image and the predefined laser dot reference colors from the RGB color space to the CIE LAB space (Pereira et al. 2019).

Thresholding and Binary Mask Creation: The ΔE value quantifies the perceptual difference between a given pixel and each reference colour. For each image pixel, the colour difference is computed with respect to all predefined laser reference colours, and the minimum ΔE value is retained as the final similarity measure. A pixel is classified as a candidate laser pixel when this

minimum value falls below a predefined threshold θ . In the implemented system, a threshold value of $\theta = 2.6$ is used to distinguish laser pixels from background pixels. This classification produces a binary segmentation mask defined as:

$$\text{Mask}(x,y) = \begin{cases} 1, & \min(\Delta E) < \theta \\ 0, & \text{otherwise} \end{cases} \quad (3.2)$$

Pixels labeled as 1 are retained for further processing as potential laser point locations, while the remaining pixels are discarded as background. This threshold-based segmentation strategy allows for precise isolation of laser regions without relying on strict geometric assumptions or prior spatial constraints.

B. Morphological Post-processing

Following CIEDE2000-based segmentation, the resulting binary mask may contain isolated noise pixels, small discontinuities, and fragmented laser regions. Figure 3.6 illustrates a typical example of these artifacts, where valid laser regions are partially disconnected and spurious foreground pixels remain after color-based segmentation. To improve the quality of the segmentation mask before localization, a morphological refinement stage is applied.

The implemented refinement procedure consists of a sequence of closing, dilation, and erosion operations. These operations are used to connect fragmented laser regions, fill small discontinuities, and suppress isolated noise pixels generated during the thresholding stage. Morphological processing is implemented using the OpenCV library with empirically selected structuring elements. The resulting refined mask contains cleaner and more coherent laser regions, thereby improving the reliability of subsequent contour extraction and bounding-box localization.

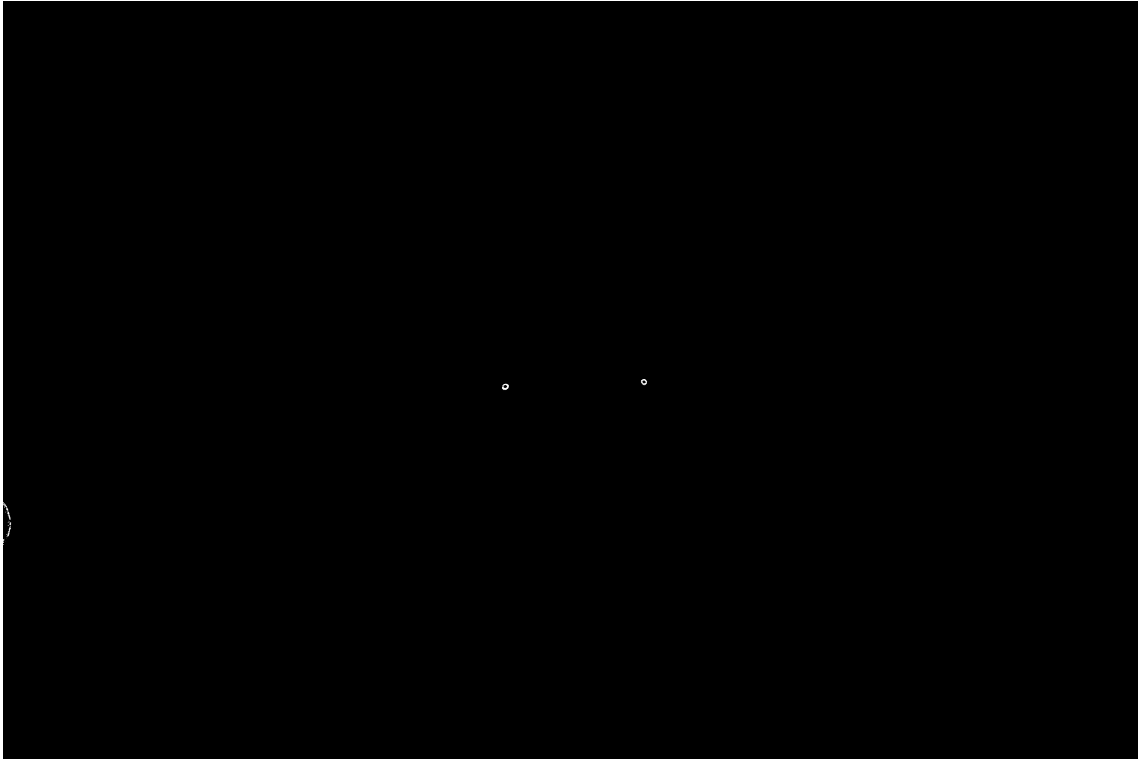


Figure 3.6: Example of artefacts present in the binary mask after CIEDE2000 thresholding, including isolated noise pixels and fragmented laser regions.

C. Object Location via Bounding Boxes

The final stage in the laser point detection pipeline involves localizing candidate laser regions by fitting bounding boxes around the segmented components. This step transforms the refined binary mask into geometrically defined detection outputs, enabling both qualitative visualization and downstream quantitative analysis. Following morphological refinement, the binary mask consists of distinct connected foreground regions corresponding to candidate laser point detections.

To identify these regions, a contour extraction algorithm is applied to detect the external boundaries of each connected foreground object. Contour extraction provides a compact representation of the shape of each detected region, enabling individual laser point candidates to be isolated from the refined binary mask. Each detected contour is assumed to represent either a valid laser point projection or a residual noise artifact that has persisted through the preceding filtering stages. For each detected contour, a bounding rectangle is computed and defined by its top-left corner coordinates together with its width and height. These bounding boxes serve as the primary object localization mechanism, providing a geometric representation of each detected laser region and the coordinates required for subsequent processing.

To reduce false positives and eliminate noise, a filtering criterion is applied based on the spatial dimensions of each bounding box. Bounding boxes whose dimensions fall outside an empirically defined range are discarded. These constraints correspond to the expected physical size of valid laser projections, ensuring that only appropriately sized detections are retained.

Each valid bounding box is projected onto the original image for visual inspection and recorded for subsequent analysis. The spatial coordinates of the detected bounding boxes are preserved in a structured representation to support geometric processing and performance evaluation. This representation enables direct comparison between automatically detected laser point locations and manually annotated ground-truth references during the evaluation stage.

Representative examples of the final laser point localization outputs produced by the proposed method are presented in Chapter 4.

3.3.2 Animal Body Segmentation

The segmentation of the animal body constitutes the second stage of the proposed framework and aims to accurately isolate the body of the cattle from the surrounding environment. Accurate segmentation is essential because the subsequent stages, including anatomical keypoint identification and morphometric measurement extraction, depend directly on the quality of the segmented animal mask.

The proposed framework employs the Segment Anything Model (SAM) for animal body segmentation. SAM is a prompt-based foundation segmentation model capable of generating high-quality object masks from simple user-defined inputs and has demonstrated strong generalization across diverse visual domains. Unlike conventional semantic segmentation models that require extensive task-specific training and large amounts of annotated data, SAM can generate segmentation masks using prompt guidance, making it suitable for livestock applications where pixel-level annotation is labor-intensive.

The central idea of the proposed segmentation approach is to use the information obtained during the laser point identification stage to guide the segmentation process. Since the detected laser markers are projected directly onto the animal body during image acquisition, they provide reliable foreground information regarding the location of the target animal. Consequently, the region located between the detected laser points belongs to the animal body and can be used to generate foreground prompts for segmentation.

For each input image, the coordinates of the detected laser points are first obtained from the laser point identification stage. Multiple foreground prompt points are then generated along the line segment connecting the detected laser markers, as illustrated in Figure 3.7. The use of multiple equally spaced prompts provides richer spatial information than a single-point prompt and improves the ability of the segmentation model to identify the target animal. Because all generated prompts are located on the animal body, they provide reliable guidance for the segmentation process.

The generated prompt points, together with the original RGB image, are provided as input to SAM. Guided by these prompts, SAM predicts multiple candidate segmentation masks corresponding to objects present in the image. The segmentation procedure consists of the following steps:

1. Detect the laser markers on the animal body.



Figure 3.7: Laser-guided prompt generation for animal segmentation. The detected laser markers are used to generate multiple equally spaced foreground prompts between the laser points. These prompts provide reliable guidance because they are located on the animal body.

2. Generate multiple foreground prompts between the detected laser points.
3. Provide the RGB image and generated prompts as input to SAM.
4. Predict candidate segmentation masks using SAM.
5. Select the mask corresponding to the target animal.
6. Convert the selected mask into a binary representation.
7. Generate visualization outputs including overlay masks and isolated animal images.

The output of this stage is a binary mask representing the segmented animal body. Additionally, overlay visualizations and cow-only images are generated for qualitative assessment. The binary mask provides an accurate representation of the animal silhouette and serves as the primary input for the subsequent anatomical keypoint identification stage. By combining laser-guided prompt generation with the segmentation capabilities of SAM, the proposed framework enables reliable animal body extraction in complex farm environments while minimizing the need for extensive manually annotated segmentation data.

3.3.3 Morphometrics Measurement

3.3.3.1 Anatomical Keypoints

Morphometric assessment of livestock animals relies on the identification of anatomically meaningful landmarks that can be consistently located across individuals. These landmarks, commonly

referred to as anatomical keypoints, are widely used in animal morphology studies to define standard body measurements such as withers height, body length, chest depth, and rump dimensions. Anatomical keypoints provide stable geometric reference locations that enable objective and repeatable estimation of body conformation traits from images.

In the proposed framework, anatomical keypoints are derived from the segmented cattle body using contour-based geometric analysis and anatomical constraints. The identified keypoints provide a structured representation of the animal anatomy and establish the relationship between the segmented body shape and the target morphometric measurements.

The anatomical representation used in this study consists of keypoints distributed across major body regions of the animal, including the head, neck, trunk, back, tail, and limbs. These landmarks correspond to anatomically relevant structures such as the Neck End, Throat End, Back Base, Back Middle, Back End, Tail Base, Front Left Paw, Front Right Paw, Back Left Paw, and Back Right Paw. Collectively, these landmarks describe the geometric structure of the cattle body and provide the reference locations required for morphometric measurement estimation.

The morphometric traits considered in this study are withers height (a_1), waist height (ANC), chest depth (a_4), body length (b_1), and rump length (b_2). These measurements are commonly used indicators of animal growth, body conformation, and overall development. Figure 3.8 illustrates the anatomical definition of the morphometric traits considered in this work.



Figure 3.8: Morphometric measurements considered in this study. a_1 : withers height; a_4 : chest depth; ANC: waist height; b_1 : body length; b_2 : rump length.

As shown in Figure 3.8, each morphometric trait is defined using anatomically meaningful reference locations that can be identified from the segmented animal body.

3.3.3.2 Anatomical Keypoint Identification

Following cattle segmentation, anatomical keypoints are identified using a heuristic-based localization approach. The objective of this stage is to automatically estimate anatomically meaningful landmarks required for morphometric measurement extraction.

The proposed approach was selected because the objective of this work is not to recover a complete skeletal pose representation, but rather to identify a limited set of anatomical landmarks

necessary for morphometric analysis. Furthermore, no large-scale dataset containing manually annotated cattle anatomical keypoints was available for training and validating a supervised pose-estimation model. Consequently, the anatomical landmarks are inferred directly from the segmented silhouette using contour geometry and anatomical knowledge of cattle morphology.

Initially, the segmented cattle body is converted into a binary mask and its external contour is extracted. The contour provides a compact geometric representation of the animal body while preserving the shape information required for anatomical analysis. Since all images are acquired from a lateral viewpoint, the contour provides a consistent representation of the dorsal body profile, ventral body profile, neck region, tail region, and limb extremities.

The extracted contour is subsequently analyzed using geometric rules derived from the spatial organization of cattle anatomy. Height-related landmarks are generated from the relationship between the dorsal body contour and the paw locations, while body-length and rump-related landmarks are derived from the relative positions of the throat, tail, and dorsal body regions. These geometric relationships allow anatomically meaningful reference points to be generated directly from the segmented silhouette.

To improve robustness, anatomical constraints derived from cattle morphology are incorporated into the localization process. Back Base, Back Middle, and Back End landmarks are constrained to the dorsal body profile, Tail Base is constrained to the posterior body region, Neck End and Throat End are constrained to the anterior body region, and paw landmarks are constrained to the lowest contour locations corresponding to the animal stance. These constraints preserve the expected spatial organization of cattle body parts and reduce anatomically inconsistent landmark detections.

The resulting anatomical landmarks provide a structured geometric representation of the cattle body that serves as the basis for morphometric measurement extraction.

The proposed anatomical keypoint localization procedure consists of the following steps:

1. Extract the binary mask generated by the segmentation stage.
2. Extract the external contour of the segmented cattle body.
3. Identify the principal anatomical regions of the animal body.
4. Generate anatomical landmarks using contour geometry and spatial relationships between body regions.
5. Apply anatomical constraints to ensure biologically plausible landmark locations.
6. Refine landmark locations and remove inconsistent detections.
7. Return the final anatomical keypoints.

3.3.3.3 Morphometric Measurement Extraction

After anatomical landmark localization, morphometric traits are estimated using predefined geometric relationships between specific landmark pairs. The identified anatomical keypoints are mapped to the target morphometric measurements according to their anatomical definitions.

The relationship between anatomical keypoints and morphometric measurements is summarized in Table 3.3.

Table 3.3: Mapping between anatomical keypoints and morphometric measurements.

Measurement	Anatomical Keypoints Used
a_1 (Withers Height)	Back Base – Midpoint of Front Paws
ANC (Waist Height)	Back End – Midpoint of Hind Paws
a_4 (Chest Depth)	Dorsal and Ventral Thoracic Reference Points
b_1 (Body Length)	Throat End – Tail Base
b_2 (Rump Length)	Tail Base – Posterior Dorsal Reference Point

3.3.3.4 Morphometric Measurement Computation and Calibration

Following anatomical keypoint identification, morphometric measurements are estimated from the generated anatomical landmarks. The proposed methodology computes measurement-specific reference points and subsequently derives morphometric traits using Euclidean distance calculations. The resulting measurements are initially expressed in image coordinates (pixels) and are later converted into real-world dimensions through laser-based calibration.

For height-related measurements, the midpoint between the corresponding paw landmarks is first computed and projected vertically through the segmented silhouette. The first foreground pixel intersecting the animal body is used as the dorsal anatomical reference point. The distance between this contour point and the corresponding paw midpoint defines the height measurement. Similarly, chest depth, body length, and rump length are estimated using anatomical reference points generated from predefined relationships between the identified landmarks. These measurement-specific reference points ensure that morphometric traits are extracted from consistent anatomical locations across different animals and images.

The midpoint of the front paws is defined as

$$P_F = \left(\frac{x_{FL} + x_{FR}}{2}, \frac{y_{FL} + y_{FR}}{2} \right), \quad (3.3)$$

where (x_{FL}, y_{FL}) and (x_{FR}, y_{FR}) denote the coordinates of the Front Left Paw and Front Right Paw landmarks, respectively.

Similarly, the midpoint of the rear paws is computed as

$$P_R = \left(\frac{x_{BL} + x_{BR}}{2}, \frac{y_{BL} + y_{BR}}{2} \right). \quad (3.4)$$

The morphometric measurements are computed as

$$a_1 = d(P_{BB}, P_F), \quad (3.5)$$

$$ANC = d(P_{BE}, P_R), \quad (3.6)$$

$$a_4 = d(P_{TE}, P_{NE}), \quad (3.7)$$

$$b_1 = d(P_{TE}, P_{TB}), \quad (3.8)$$

and

$$b_2 = d(P_{TB}, P_M), \quad (3.9)$$

where

$$P_M = \left(\frac{x_{BE} + x_{BM}}{2}, \frac{y_{BE} + y_{BM}}{2} \right) \quad (3.10)$$

represents the midpoint between the Back End and Back Middle landmarks.

The Euclidean distance function is defined as

$$d(P_i, P_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}. \quad (3.11)$$

The resulting measurements are initially expressed in pixels. To obtain real-world dimensions, a calibration procedure based on the dual-laser acquisition system is employed. During image acquisition, two laser points separated by a known physical distance of 32 cm are projected onto the animal body. Following laser point detection, the distance between the detected laser projections is measured in image coordinates and used to estimate a pixel-to-centimeter conversion factor.

$$S = \frac{D_{real}}{D_{pixel}}, \quad (3.12)$$

where $D_{real} = 32$ cm and D_{pixel} denotes the distance between the detected laser points in image coordinates.

The final morphometric measurement expressed in centimeters is obtained as

$$M_{cm} = M_{pixel} \times S. \quad (3.13)$$

Consequently, all morphometric traits are reported in centimeters and can be directly compared with manually acquired ground-truth measurements during the quantitative evaluation stage.

3.4 Comparison with other models

To rigorously evaluate the effectiveness of the proposed segmentation approach based on the Segment Anything Model (SAM), a comprehensive benchmarking study was conducted using three widely adopted deep learning architectures: U-Net, DeepLabV3+, and Feature Pyramid Network (FPN). These models were selected due to their strong performance in semantic segmentation tasks and their representation of diverse architectural design principles.

3.4.1 Dataset and Experimental Setup

All models were trained and evaluated using the same dataset of cattle images described in Section 3.2. The dataset consists of high-resolution RGB images with corresponding ground-truth segmentation masks representing the animal body. No additional data or external datasets were introduced during training or evaluation, ensuring a controlled and consistent experimental environment.

3.4.2 Preprocessing and Data Consistency

All benchmark models were trained using identical preprocessing steps to ensure fairness in comparison. These steps include image normalization, resizing where necessary to fit model input requirements, and data augmentation techniques such as horizontal flipping and rotation. By maintaining a consistent preprocessing pipeline, performance differences can be attributed solely to model architecture rather than data handling variations.

These benchmark models were implemented under the same experimental conditions as the proposed SAM-based method to ensure a fair and consistent comparison.

3.5 Evaluation Metrics

To quantitatively assess the proposed system, evaluation metrics were defined for laser point detection, segmentation quality, and morphometric measurement accuracy. These metrics were computed using predicted outputs and their corresponding ground-truth annotations.

3.5.1 Laser Point Detection Evaluation Metrics

The performance of the laser point detection module was evaluated by comparing the automatically detected laser-point locations with manually annotated ground-truth references. The evaluation was performed for both the left and right laser projections and included region-based and point-based localization metrics.

The evaluation dataset consisted of manually annotated images containing the locations of the projected laser points. For each image, the proposed detection algorithm generated bounding boxes surrounding the detected laser regions. The coordinates and dimensions of the detected bounding boxes were subsequently compared with the corresponding ground-truth bounding boxes.

Intersection over Union (IoU): The overlap between the predicted and ground-truth laser-point regions was quantified using the Intersection over Union (IoU) metric. IoU measures the degree of agreement between the detected and reference bounding boxes and is defined as:

$$IoU = \frac{A_{intersection}}{A_{union}} \quad (3.14)$$

where $A_{intersection}$ represents the overlapping area between the predicted and ground-truth bounding boxes and A_{union} denotes the total area covered by both bounding boxes.

Higher IoU values indicate better localization performance, with an IoU value of 1 corresponding to perfect overlap and an IoU value of 0 indicating no overlap.

Laser Center Localization Error: In addition to region overlap, the positional accuracy of the detected laser points was evaluated using the Euclidean distance between the detected laser-point center and the corresponding manually annotated ground-truth center.

The localization error was computed as:

$$d = \sqrt{(x_{pred} - x_{gt})^2 + (y_{pred} - y_{gt})^2} \quad (3.15)$$

where (x_{pred}, y_{pred}) corresponds to the coordinates of the laser-point center detected and (x_{gt}, y_{gt}) corresponds to the coordinates of the manually annotated ground-truth center.

The Euclidean distance provides a direct measure of the localization error in pixel units. Smaller distance values indicate more accurate detection of the laser-point position.

Descriptive Statistics: For both IoU and Euclidean distance error metrics, descriptive statistics were computed across the evaluation dataset. These statistics include the mean, median, minimum, maximum, and standard deviation.

The mean value provides the average performance for all the images evaluated, while the median value represents the central tendency of the distribution. The minimum and maximum values indicate the best and worst observed performances, respectively. The standard deviation measures the variability of the detection accuracy across different images and acquisition conditions.

The evaluation results were computed separately for the left and right laser projections to assess the consistency and robustness of the proposed laser point detection approach.

3.5.2 Segmentation Metrics

Segmentation performance was evaluated using standard pixel-wise classification metrics. Let TP , FP , TN , and FN denote true positives, false positives, true negatives, and false negatives, respectively.

$$IoU = \frac{TP}{TP + FP + FN} \quad (3.16)$$

$$Dice = \frac{2TP}{2TP + FP + FN} \quad (3.17)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.18)$$

$$Precision = \frac{TP}{TP + FP} \quad (3.19)$$

$$Recall = \frac{TP}{TP + FN} \quad (3.20)$$

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (3.21)$$

3.5.3 Measurement Metrics

The accuracy of predicted morphometric measurements was evaluated using regression-based error metrics. For predicted measurement M_{pred} and ground truth M_{gt} , the absolute error is defined as:

$$AE = |M_{pred} - M_{gt}| \quad (3.22)$$

Aggregate metrics include:

$$MAE = \frac{1}{N} \sum |M_{pred} - M_{gt}| \quad (3.23)$$

$$MSE = \frac{1}{N} \sum (M_{pred} - M_{gt})^2 \quad (3.24)$$

$$RMSE = \sqrt{MSE} \quad (3.25)$$

$$MAPE = \frac{1}{N} \sum \left| \frac{M_{pred} - M_{gt}}{M_{gt}} \right| \quad (3.26)$$

In addition, Median Absolute Error, Maximum Error, and the coefficient of determination (R^2) were computed to further evaluate prediction performance.

Chapter 4

Results

This chapter presents the experimental results obtained from the proposed bovine morphometric analysis framework. The evaluation encompasses the four principal stages of the system: laser point detection, cattle segmentation, anatomical keypoint localization, and morphometric measurement estimation.

The analysis begins with laser point detection, which provides the spatial reference required for scale calibration. Segmentation results are then presented to assess the ability of the proposed approach to accurately isolate cattle from complex farm environments. Subsequently, the performance of the anatomical keypoint localization module is evaluated through qualitative and quantitative analyses. Finally, the estimated morphometric measurements are compared with ground-truth measurements to assess the accuracy and reliability of the proposed measurement estimation framework.

Results are presented using both qualitative visual examples and quantitative performance metrics derived from the evaluation dataset. Together, these results provide a comprehensive assessment of the effectiveness of the proposed framework under real livestock farm conditions.

4.1 Laser Point Detection Results

4.1.1 Quantitative Results

The quantitative performance of the laser point detection module was evaluated using two complementary metrics: Intersection over Union (IoU) and Euclidean distance error. IoU was used to assess the overlap between the detected and manually annotated laser-point bounding boxes, whereas Euclidean distance error was used to evaluate the positional accuracy of the detected laser-point centers.

Table 4.1 presents the IoU statistics obtained for the left and right laser projections across the evaluation dataset.

Table 4.1: Intersection over Union (IoU) results for laser point detection.

Statistic	Left Laser	Right Laser
Mean IoU	0.838	0.829
Median IoU	0.856	0.840
Standard Deviation	0.089	0.092
Minimum IoU	0.436	0.410
Maximum IoU	1.000	0.960

In addition to IoU, the localization accuracy of the detected laser-point centers was evaluated using the Euclidean distance error. The corresponding statistics are summarized in Table 4.2.

Table 4.2: Euclidean distance error of laser-point center localization.

Statistic	Left Laser	Right Laser
Mean Error (pixels)	1.10	1.40
Median Error (pixels)	1.12	1.12
Standard Deviation	0.57	0.88
Minimum Error (pixels)	0.00	0.50
Maximum Error (pixels)	2.55	6.02

The quantitative results presented in Tables 4.1 and 4.2 indicate that the proposed laser point detection approach achieved accurate localization of the projected laser markers. The high IoU values demonstrate substantial overlap between the detected and ground-truth laser regions, while the low Euclidean distance errors indicate accurate localization of the laser-point centers. These results confirm that the detected laser points provide a reliable basis for the subsequent scale calibration and morphometric measurement estimation stages.

4.1.2 Visual Results

Representative examples of the intermediate and final outputs produced by the proposed laser point detection pipeline are presented in Figures 4.1, 4.2, and 4.3. These figures illustrate the progression from color-based segmentation to the final localization of the projected laser points.

Figure 4.1 presents the binary mask generated after the CIEDE2000 thresholding stage. The thresholding procedure successfully isolates candidate laser regions from the background based on their color similarity to the predefined laser reference colors. Although the laser regions are clearly identified, small isolated foreground pixels and fragmented regions remain visible due to image noise, illumination variations, and surface reflectance effects.

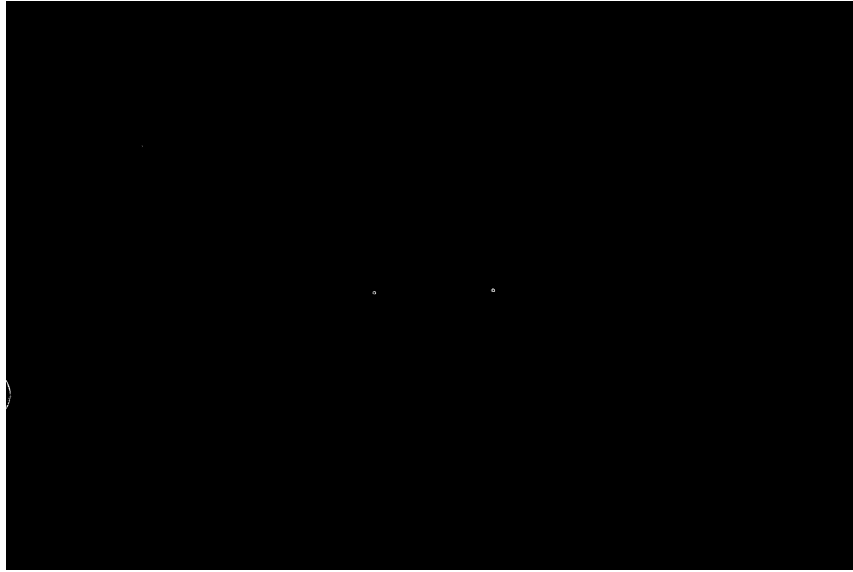


Figure 4.1: Binary mask obtained after CIEDE2000 thresholding. Candidate laser regions are separated from the background based on color similarity.

Figure 4.2 shows the result obtained after morphological refinement. The application of closing, dilation, and erosion operations improves the connectivity of the detected laser regions and suppresses isolated noise pixels. Consequently, the segmented laser regions become more compact and coherent, providing a cleaner representation of the projected laser points for the subsequent localization stage.

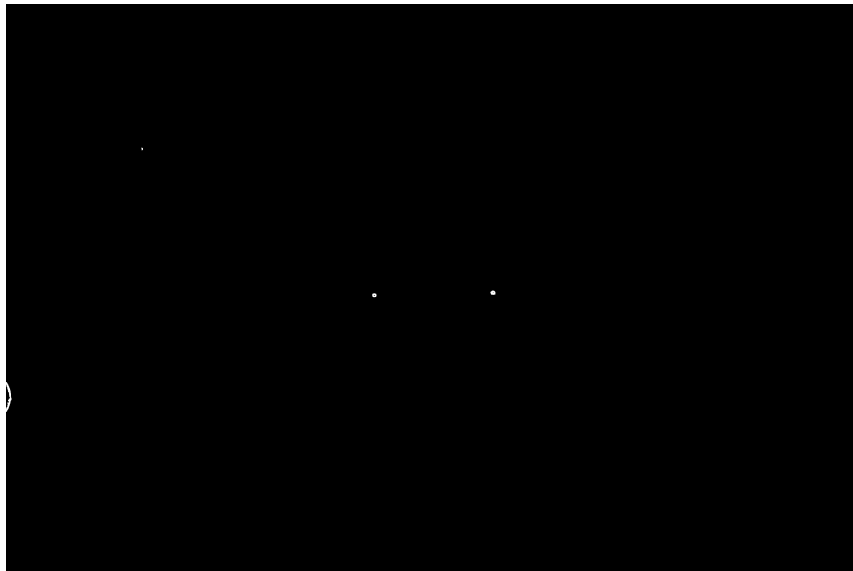


Figure 4.2: Refined binary mask after morphological processing. Noise pixels are reduced and the detected laser regions become more compact and connected.

The final localization results are presented in Figure 4.3. Following contour extraction and

bounding-box fitting, the detected laser regions are accurately localized and enclosed within bounding boxes. The resulting detections provide the spatial coordinates required for scale calibration and subsequent morphometric measurement estimation.



Figure 4.3: Examples of final laser-point localization obtained through contour extraction and bounding-box detection. The detected laser regions are enclosed within bounding boxes and used as spatial references for scale calibration.

Overall, the qualitative results demonstrate the effectiveness of the proposed laser point detection pipeline. The sequential processing stages progressively refine the detected laser regions, resulting in accurate localization of the projected laser markers under varying image acquisition conditions.

4.2 Segmentation Results of the Proposed SAM-Based Approach

Cattle body segmentation constitutes the second stage of the proposed framework and provides the animal regions required for anatomical keypoint localization and morphometric measurement estimation. Segmentation masks were generated using the Segment Anything Model (SAM), producing foreground representations of the cattle while suppressing background structures present in the farm environment.

4.2.1 Qualitative Results

The segmentation outputs were examined through qualitative visual analysis using representative samples from the dataset. The generated segmentation overlays, binary masks, and extracted cattle regions are illustrated in Figures 4.4, 4.5, and 4.6, respectively.

Figure 4.4 shows representative segmentation overlays generated by SAM. The highlighted regions correspond to the cattle body identified by the segmentation model. The examples include images acquired under different environmental conditions, including variations in illumination, background composition, and animal pose.



Figure 4.4: Representative cattle segmentation overlays generated using the Segment Anything Model (SAM). The highlighted regions correspond to the cattle body identified by the segmentation model. The examples demonstrate the segmentation outputs obtained under different acquisition conditions and background environments.

The corresponding binary masks are presented in Figure 4.5. The masks provide a simplified representation of the segmented cattle regions and suppress most background structures present in the original images. The overall body shape of the animals is preserved, providing the foreground regions required for subsequent processing.



Figure 4.5: Representative binary segmentation masks generated using SAM. White pixels correspond to the detected cattle regions, while black pixels represent the background.

Figure 4.6 presents the cattle regions obtained after applying the generated masks to the original images. The segmented outputs retain the visible anatomical structure of the animals while removing most surrounding environmental elements. These segmented cattle regions are subsequently used for anatomical keypoint localization and morphometric measurement estimation.

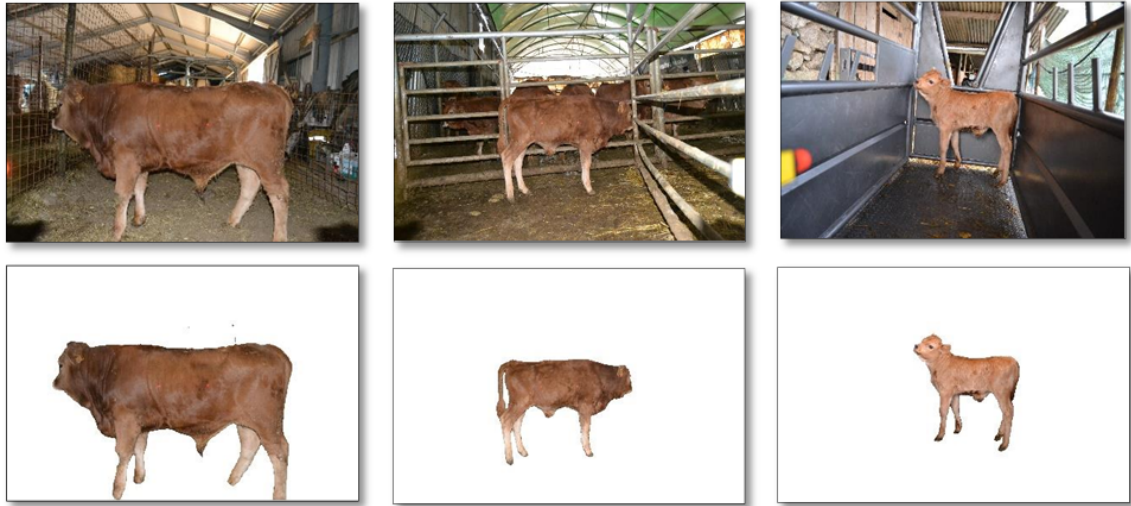


Figure 4.6: Representative segmentation overlays generated using SAM. The highlighted regions correspond to the cattle body identified by the segmentation model.

The examples presented in Figures 4.4, 4.5, and 4.6 illustrate the segmentation outputs generated by SAM for representative images from the dataset. The resulting masks provide the cattle regions required for the subsequent stages of the proposed framework.

Since manually annotated segmentation masks were not available for the cattle dataset, the segmentation quality of the proposed SAM-based approach was assessed through qualitative visual inspection rather than quantitative segmentation metrics. The visual analysis demonstrated that SAM successfully identified the cattle body while suppressing most background structures present in the livestock environment. The generated masks preserved the overall body shape and major anatomical contours required for subsequent anatomical keypoint localization and morphometric measurement extraction. Furthermore, the segmentation outputs remained consistent under different acquisition conditions, including variations in illumination, animal pose, and background complexity. Based on this qualitative assessment, the masks generated by SAM were considered sufficiently accurate for the subsequent stages of the proposed framework. Consequently, the SAM-generated masks were adopted as pseudo-ground-truth annotations for training and evaluating the benchmark segmentation models.

4.2.2 Results of Benchmark Models

To further investigate the effectiveness of the proposed SAM-based segmentation approach, three benchmark semantic segmentation models, namely U-Net, DeepLabV3+, and Feature Pyramid Network (FPN), were evaluated.

Since manually annotated segmentation masks were not available for the cattle dataset, the segmentation masks generated by SAM were adopted as pseudo-ground-truth annotations. Consequently, the benchmark models were trained and evaluated using the SAM-generated masks as target labels. Therefore, the purpose of this comparison was not to determine whether the benchmark models outperform the proposed SAM-based approach, but rather to evaluate the degree of agreement between the benchmark model predictions and the segmentation masks generated by SAM.

The benchmark models demonstrated the ability to reproduce the segmentation outputs generated by SAM with varying levels of similarity. U-Net benefited from its encoder–decoder architecture and skip connections, DeepLabV3+ leveraged atrous convolution for contextual feature extraction, and FPN provided effective multi-scale feature representation. These models were subsequently analyzed through both qualitative and quantitative evaluation.

4.2.3 Visual Results of Benchmark Models

The qualitative segmentation results obtained using the benchmark models, namely U-Net, FPN, and DeepLabV3+, are presented in Figures 4.7, 4.8, and 4.9, respectively.

As illustrated in Figures 4.7–4.9, all benchmark models were able to identify the main cattle body region and generate segmentation masks that were generally consistent with the masks produced by the proposed SAM-based approach. However, differences can be observed in boundary preservation, contour accuracy, and background suppression.

U-Net generally produced accurate segmentation masks with well-preserved body contours. DeepLabV3+ achieved comparable results but occasionally introduced background noise and incomplete boundaries in challenging regions affected by illumination variations and complex backgrounds. Similarly, FPN demonstrated effective cattle segmentation but showed limitations in preserving fine contour details and accurately delineating certain boundary regions.

Overall, the visual assessment indicates that the benchmark models successfully captured the principal cattle body structure, although differences were observed in contour preservation and boundary accuracy.

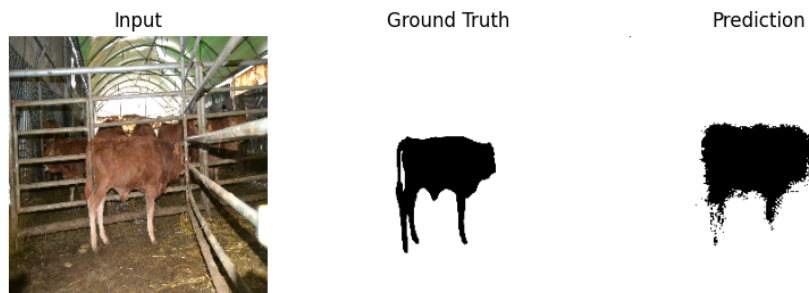


Figure 4.7: Representative segmentation results obtained using the U-Net model. From left to right: input image, SAM-generated pseudo-ground-truth mask, and predicted segmentation mask.

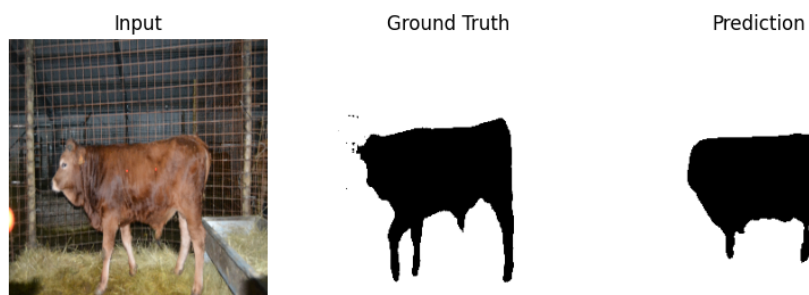


Figure 4.8: Representative segmentation results obtained using the FPN model. From left to right: input image, SAM-generated pseudo-ground-truth mask, and predicted segmentation mask.

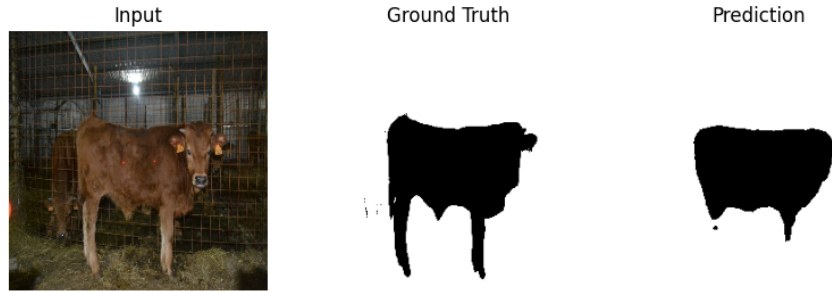


Figure 4.9: Representative segmentation results obtained using the DeepLabV3+ model. From left to right: input image, SAM-generated pseudo-ground-truth mask, and predicted segmentation mask.

4.2.4 Quantitative Results of Benchmark Models

The quantitative performance of the benchmark segmentation models was evaluated using Intersection over Union (IoU), Dice Score, Accuracy, Precision, Recall, and F1-score. Since manually annotated segmentation masks were unavailable, the SAM-generated masks were adopted as pseudo-ground-truth annotations. Consequently, the reported metrics quantify the similarity between the benchmark model predictions and the segmentation masks generated by the proposed SAM-based approach.

The quantitative evaluation results are summarized in Table 4.3.

Table 4.3: Quantitative performance comparison of benchmark segmentation models.

Model	IoU	Dice	Accuracy	Precision	Recall	F1-score
U-Net	0.8991	0.9467	0.9099	0.9634	0.9310	0.9467
FPN	0.8840	0.9380	0.8967	0.9651	0.9126	0.9380
DeepLabV3+	0.8916	0.9426	0.9037	0.9667	0.9203	0.9426

As shown in Table 4.3, all benchmark models achieved high agreement with the SAM-generated masks, indicating that they successfully reproduced the principal segmentation characteristics captured by the proposed SAM-based approach. Among the evaluated models, U-Net achieved the highest overall agreement, while DeepLabV3+ demonstrated comparable performance. FPN also produced strong results, although its agreement scores were slightly lower.

These findings suggest that the segmentation patterns generated by the proposed SAM-based approach can be effectively learned and reproduced by conventional semantic segmentation architectures. However, because SAM served as the source of the pseudo-ground-truth annotations, the

reported metrics should be interpreted as measures of agreement with the SAM-generated masks rather than absolute segmentation accuracy.

4.3 Morphological Measurements Results

Automatic cattle morphometric analysis was performed using the proposed segmentation, anatomical keypoint identification, contour analysis, and laser-based calibration framework. The extracted anatomical landmarks were used to estimate five morphometric traits directly from RGB images: withers height ($a1$), height at tail insertion (ANC), chest depth ($a4$), body length ($b1$), and rump length ($b2$).

The morphometric measurements were computed from the automatically detected anatomical reference points together with the laser-based pixel-to-centimeter calibration procedure described in Chapter 3. The detected laser points provided a reliable spatial reference for converting image-based distances into real-world physical measurements.

4.3.1 Visual Analysis of Morphological Measurements

Figure 4.10 presents representative contour extraction results obtained under different livestock farm environmental conditions. The extracted contours preserved the geometric structure of the cattle body while minimizing background interference, enabling reliable anatomical landmark localization for morphometric analysis.

The segmentation framework effectively isolated the animal body from surrounding background regions and maintained stable contour generation despite variations in animal posture, body orientation, illumination conditions, fur appearance, and environmental background complexity. The detected laser reference points also remained visible in most samples, supporting consistent spatial calibration during measurement estimation.

Figure 4.11 illustrates examples of generated morphometric measurement references, including withers height ($a1$), waist height (ANC), chest depth ($a4$), body length ($b1$), and rump length ($b2$). The generated anatomical landmarks closely followed the cattle body structure, demonstrating the effectiveness of the proposed contour-based anatomical measurement framework.

Figure 4.12 compares manually annotated ground-truth (GT) anatomical landmarks with the generated point visualization outputs. The generated landmark points closely matched the manually annotated reference points across different cattle samples, indicating reliable anatomical landmark localization under varying farm conditions. Figure 4.13 presents a representative example of the output generated by the anatomical keypoint identification stage. Following animal segmentation, a set of anatomical landmarks is automatically localized on the cattle body. These landmarks describe relevant anatomical structures, including the head, neck, dorsal profile, tail base, and limb extremities. The geometric relationships established between the identified key-points provide the foundation for the subsequent distance calculation process and the estimation of morphometric traits.

Although the framework produced consistent outputs in most cases, minor inaccuracies were observed in some images affected by partial occlusion, shadow interference, incomplete body visibility, and irregular animal posture. These factors occasionally introduced contour discontinuities and slight landmark localization errors, which affected the final measurement estimation.

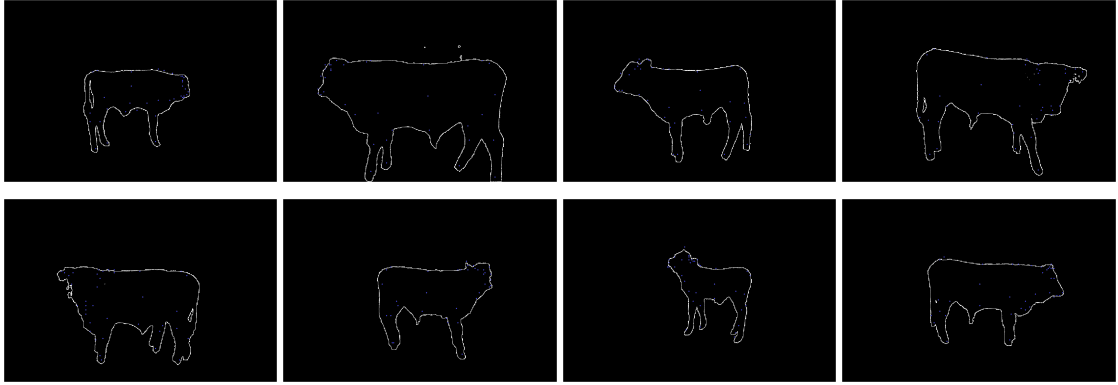


Figure 4.10: Representative contour extraction results generated from the cattle image dataset. The extracted contours preserved the geometric structure of the animal body while minimizing background interference, enabling reliable anatomical landmark localization for morphometric analysis under varying environmental conditions.



(a)



(b)



(c)



(d)

Figure 4.11: Representative examples of morphometric distance calculation obtained from the automatically identified anatomical keypoints. The generated measurement references illustrate the estimation of withers height (a_1), waist height (ANC), chest depth (a_4), body length (b_1), and rump length (b_2) from segmented cattle images.



(a) Reference anatomical keypoints.



(b) Automatically identified anatomical keypoints.



(c) Reference anatomical keypoints.

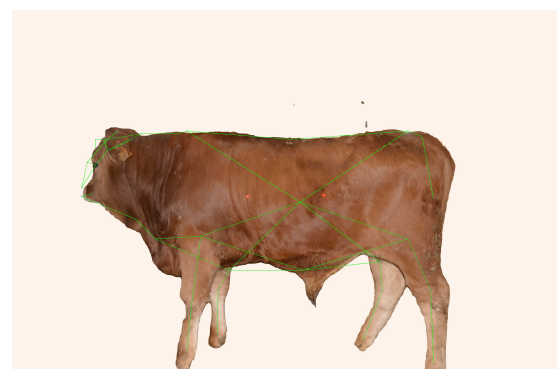


(d) Automatically identified anatomical keypoints.

Figure 4.12: Representative results of Stage 3 of the proposed framework: anatomical keypoint identification and distance calculation. The figure compares manually annotated reference landmarks with the anatomical keypoints automatically generated by the proposed method. The identified landmarks provide the geometric basis for estimating morphometric traits, including withers height (a_1), waist height (ANC), chest depth (a_4), body length (b_1), and rump length (b_2).



(a)



(b)

Figure 4.13: Representative outputs of the anatomical keypoint identification stage. The automatically generated landmarks provide a structured geometric representation of the cattle body and establish the reference points required for morphometric distance calculation. These landmarks are subsequently used to estimate morphometric traits, including withers height (a_1), waist height (ANC), chest depth (a_4), body length (b_1), and rump length (b_2).

4.3.2 Quantitative Morphometric Measurement Evaluation

The quantitative evaluation assesses the ability of the proposed framework to estimate morphometric measurements from automatically identified anatomical keypoints. The predicted measurements were compared with the corresponding manually acquired ground-truth measurements for the five morphometric traits considered in this study: withers height (a_1), height at tail insertion (ANC), chest depth (a_4), body length (b_1), and rump length (b_2). The evaluation was performed using the regression metrics described in Section 3.5.3.

To provide a comprehensive analysis, the quantitative results are presented in two complementary forms. First, representative prediction examples are provided to illustrate the agreement between the automatically estimated and manually measured morphometric traits for individual animals. Second, the overall performance of the proposed framework is summarized using consolidated evaluation metrics computed over the complete dataset. This presentation provides both detailed insight into representative prediction cases and an overall assessment of the framework's accuracy and robustness.

Table 4.4: Ground-truth and predicted morphometric measurements for six representative cattle.

Animal	a1		ANC		a4		b1		b2		Total Err.
	GT	Pred	GT	Pred	GT	Pred	GT	Pred	GT	Pred	
195	80.5	81.10	85.5	85.98	38.5	40.00	99.5	101.81	30.0	22.84	12.05
52	101.5	99.13	109.5	109.10	45.0	47.10	112.0	118.45	35.0	33.35	12.97
84	85.0	85.19	94.0	94.33	41.5	40.66	110.0	103.14	34.0	26.88	15.34
208	73.5	73.30	76.5	82.41	34.0	28.52	92.0	89.99	26.0	23.95	15.65
155	83.0	86.90	93.0	101.49	46.0	46.94	119.0	119.85	35.0	31.53	17.65
73	75.0	80.55	79.0	82.60	37.0	35.84	96.0	92.71	28.0	22.38	19.22

Table 4.4 presents representative examples of the morphometric measurements estimated by the proposed framework. The selected examples demonstrate that the automatically estimated measurements generally follow the corresponding ground-truth values, indicating that the proposed anatomical keypoint identification and measurement strategy is capable of accurately estimating morphometric traits from RGB images. Although the prediction accuracy varies among individual animals, the overall measurement trends remain consistent across the evaluated traits. The observed discrepancies are primarily associated with challenging imaging conditions, including variations in animal posture, contour shape, partial occlusions, and the difficulty of precisely localizing anatomical landmarks. These representative cases illustrate both the strengths of the proposed framework and the types of prediction errors encountered under practical farm conditions.

Table 4.5: Overall quantitative performance of the proposed morphometric measurement framework evaluated on the complete dataset of 66 cattle.

Measurement	MAE (cm)	RMSE (cm)	MAPE (%)	Median AE (cm)	Max Error (cm)
<i>a</i> 1 (Withers Height)	11.09	13.85	11.72	9.93	42.25
<i>ANC</i> (Height at Tail Insertion)	12.70	14.85	12.58	11.95	37.99
<i>a</i> 4 (Chest Depth)	6.56	8.68	14.83	4.87	24.54
<i>b</i> 1 (Body Length)	11.31	14.41	10.07	8.54	45.53
<i>b</i> 2 (Rump Length)	6.86	7.81	18.82	6.46	14.36

Table 4.5 summarizes the quantitative performance of the proposed framework over the complete evaluation dataset consisting of 66 cattle. Unlike the representative examples shown in Table 4.4, these results provide a global assessment of the framework’s prediction accuracy across all evaluated animals. The consolidated metrics demonstrate that the proposed methodology consistently estimates the selected morphometric traits with acceptable accuracy under practical farm conditions.

The prediction performance varies among the evaluated measurements, reflecting differences in anatomical complexity and landmark localization difficulty. Measurements defined by anatomically distinct landmarks generally exhibit more consistent predictions, whereas measurements depending on less prominent anatomical reference points are more susceptible to localization uncertainty. Similarly, morphometric traits with relatively small physical dimensions are inherently more sensitive to minor localization errors, resulting in comparatively higher relative errors despite maintaining low absolute errors. The presence of occasional larger prediction errors further suggests that challenging samples, rather than systematic deficiencies of the proposed framework, are the primary source of performance variability.

Overall, the quantitative evaluation demonstrates that the proposed framework provides reliable and robust morphometric measurements across the complete evaluation dataset. The results confirm the effectiveness of integrating laser-guided calibration, animal segmentation, and anatomical keypoint identification into a unified computer vision framework for automated cattle morphometric analysis.

Chapter 5

Conclusions

This dissertation addressed the challenge of developing an automated and non-invasive framework for bovine morphometric analysis using computer vision techniques. Traditional livestock measurement procedures rely heavily on manual operations that are time-consuming, labor-intensive, and susceptible to human error. To overcome these limitations, a framework integrating laser point detection, laser-based scale calibration, animal segmentation, pose estimation, anatomical keypoint identification, and morphometric measurement extraction was proposed and evaluated.

5.1 Final Discussion

The primary objective of this dissertation was to investigate the feasibility of automatically extracting cattle morphometric measurements from monocular RGB images acquired under practical farm conditions. To achieve this objective, a computer vision-based framework integrating laser point detection, laser-based calibration, animal segmentation, anatomical keypoint identification, and morphometric measurement estimation was proposed and experimentally evaluated. The obtained results demonstrate that the proposed framework successfully estimates important morphometric traits while providing a non-invasive and cost-effective alternative to conventional manual measurement procedures.

The experimental evaluation confirms the effectiveness of the proposed framework as an integrated solution for automated cattle morphometric analysis. The laser point detection stage provided a reliable spatial reference for converting image measurements into real-world dimensions, while the animal segmentation stage effectively isolated the cattle body from complex farm backgrounds. These outputs formed the basis for the anatomical keypoint identification stage, enabling the automatic estimation of morphometric measurements. The successful integration of these three stages demonstrates that the proposed methodology can operate consistently under realistic livestock farming conditions.

The qualitative and quantitative results demonstrate that the proposed framework is capable of accurately estimating the selected morphometric traits. The representative examples show close agreement between the automatically estimated measurements and the corresponding manually

acquired ground-truth values, while the consolidated evaluation over the complete dataset confirms the overall robustness and consistency of the proposed methodology. These findings indicate that the integration of laser-guided calibration, segmentation, and anatomical keypoint identification provides a reliable basis for automatic morphometric measurement using monocular RGB images.

Although the overall performance of the proposed framework is encouraging, the experimental evaluation also highlights several challenges associated with image-based morphometric analysis. The prediction accuracy is not uniform across all morphometric traits because each measurement depends on the accurate localization of specific anatomical landmarks. Measurements defined by clearly distinguishable anatomical structures generally produced more consistent estimates, whereas measurements relying on less prominent landmarks exhibited greater variability. Furthermore, perspective distortion, animal pose variation, contour irregularities, partial occlusions, and illumination changes introduce additional uncertainty during anatomical keypoint localization, which subsequently propagates to the final morphometric measurements.

The obtained results suggest that anatomical keypoint identification represents the most challenging stage of the proposed framework. Since the morphometric measurements are directly computed from the detected anatomical landmarks, small localization errors may propagate throughout the measurement estimation process. Nevertheless, the overall experimental results indicate that these errors do not compromise the general applicability of the proposed framework, which consistently produced reliable morphometric measurements across the evaluation dataset.

From a practical perspective, the proposed framework contributes to precision livestock farming by demonstrating that computer vision techniques can automate cattle morphometric measurement without requiring physical contact with the animals. Compared with traditional manual measurement procedures, the proposed methodology reduces human intervention, improves operational efficiency, and provides a scalable solution for livestock growth monitoring, breeding programmes, and data-driven farm management.

Overall, the objectives established for this dissertation were successfully achieved. The proposed framework demonstrates the feasibility of integrating laser point detection, laser-based calibration, animal segmentation, anatomical keypoint identification, and morphometric measurement estimation into a unified computer vision pipeline for automated cattle morphometric analysis. Although the experimental evaluation identified limitations associated with monocular image analysis and anatomical landmark localization, the obtained results demonstrate that the proposed methodology provides a solid foundation for future research on intelligent livestock monitoring systems and supports the continued adoption of artificial intelligence and computer vision technologies in precision livestock farming.

5.2 Future Work

Although the proposed framework demonstrated promising results for automated cattle morphometric measurement from monocular RGB images, several opportunities remain to further improve its accuracy, robustness, and practical applicability.

The principal limitation of the proposed framework lies in the anatomical keypoint identification stage. Since the final morphometric measurements are directly computed from the detected anatomical landmarks, even small localization inaccuracies propagate to the subsequent measurement estimation process and affect the overall prediction accuracy. Future work should therefore investigate dedicated deep learning-based anatomical keypoint detection models trained using manually annotated cattle anatomical landmarks. Such approaches are expected to improve landmark localization accuracy, reduce measurement errors, and enhance the robustness and reliability of the proposed framework.

Another promising research direction is the extension of the proposed framework from two-dimensional (2D) image analysis to three-dimensional (3D) cattle reconstruction. Although the proposed methodology successfully estimates morphometric measurements from monocular RGB images using laser-based calibration, the absence of depth information makes the measurements susceptible to perspective distortion, animal pose variation, and partial occlusions. Future research could investigate stereo vision, RGB-D cameras, structured-light sensors, or multi-view image reconstruction techniques to generate accurate three-dimensional representations of the animal body. Incorporating depth information would improve anatomical landmark localization, increase measurement accuracy, and enable the estimation of additional morphometric traits, body volume, and live body weight, thereby extending the applicability of the proposed framework for precision livestock farming.

Future studies should also evaluate the proposed framework using larger and more diverse datasets comprising different cattle breeds, age groups, body conformations, and farming environments. Increasing the diversity of the dataset would improve the generalization capability of the proposed methodology and provide a more comprehensive assessment of its robustness under practical livestock production conditions.

Another important research direction involves improving the computational efficiency of the proposed framework for real-time deployment. Optimizing the segmentation and anatomical keypoint identification stages through lightweight deep learning models would facilitate implementation on mobile devices or edge-computing platforms, enabling continuous and automated livestock monitoring directly on farms without requiring high-performance computing resources.

In addition, the current study focused on the automatic estimation of five representative morphometric traits. Future research may extend the proposed framework to estimate additional anatomical measurements and investigate its applicability to other livestock species. Expanding the range of automatically estimated morphometric traits would provide a more comprehensive assessment of animal body conformation for breeding, growth monitoring, and production management applications.

Finally, the automatically estimated morphometric measurements could be integrated with intelligent machine learning models for body weight prediction, body condition scoring, health monitoring, and growth assessment. Combining automated image-based morphometric analysis with decision-support systems would further contribute to precision livestock farming by enabling continuous, objective, and non-invasive monitoring of livestock throughout their production cycle.

Overall, these future research directions have the potential to improve the accuracy, robustness, scalability, and practical applicability of the proposed framework, contributing to the development of fully automated and intelligent livestock monitoring systems for precision livestock farming.

References

- Andrew, William, Colin Greatwood, and Tilo Burghardt. 2017. “Visual Localisation and Individual Identification of Holstein Friesian Cattle via Deep Learning.” In *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2850–2859.
- Barbal, Francesco, Christoph Bahr, and Daniel Berckmans. 2023. “Keypoint detection for locomotion analysis in dairy cows.” *Animals* 13 (3): 412. <https://doi.org/10.3390/ani13030412>.
- Bati, A., and S. Khan. 2025. “Sheep pose estimation using convolutional neural networks and object detection.” *Computers and Electronics in Agriculture* 210:124–136.
- Bommasani, Rishi, et al. 2021. “On the Opportunities and Risks of Foundation Models.” *arXiv preprint arXiv:2108.07258*.
- Cangar, Özlem, Thomas Leroy, Marcella Guarino, Erik Vranken, Richard Fallon, John Lenehan, John Mee, and Daniel Berckmans. 2008. “Automatic estimation of cattle weight using image analysis.” *Computers and Electronics in Agriculture* 64 (2): 166–173. <https://doi.org/10.1016/j.compag.2008.05.003>.
- Cao, Jinkun, Hongyang Tang, Hao-Shu Fang, Xiaoyong Shen, Cewu Lu, and Yu-Wing Tai. 2019. “Cross-domain adaptation for animal pose estimation.” In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9498–9507.
- Chen, Liang-Chieh, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. 2018. “Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation.” In *European Conference on Computer Vision (ECCV)*, 801–818. Springer.
- Diwan, T., G. Anirudh, J. V. Tembhurne, et al. 2022. “Object Detection Using YOLO: Challenges, Architectural Successes, Datasets, Performance Metrics and Application Areas.” *Sensors* 22 (19): 7529.
- Dong, Yongsheng, Chongchong Mao, Lintao Zheng, Qingtao Wu, Mingchuan Zhang, and Xuelong Li. 2026. “AFPN: Alignment Feature Pyramid Network for Real-Time Semantic Segmentation.” *Pattern Recognition* 170:112019.
- Dosovitskiy, Alexey, et al. 2021. “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale.” In *ICLR*.
- Feng, Tao, Yangyang Guo, Xiaoping Huang, and Yongliang Qiao. 2023. “Cattle Target Segmentation Method in Multi-Scenes Using Improved DeepLabV3+ Method.” *Animals* 13 (15): 2521.
- Hansen, M. F., M. L. Smith, L. N. Smith, M. G. Salter, E. M. Baxter, M. Farish, and B. Grieve. 2018. “Towards automated lameness detection in dairy cattle using deep learning-based keypoint localization.” *Computers and Electronics in Agriculture* 153:1–14.

- Hansen, Mikkel F., Martin Smith, and Lindsay Smith. 2021. "Facial landmark detection for dairy cow emotion and welfare assessment." *Animals* 11 (9): 2608. <https://doi.org/10.3390/ani11092608>.
- Hao, Shijie, Yuan Zhou, and Yanrong Guo. 2020. "A Brief Survey on Semantic Segmentation with Deep Learning." *Neurocomputing* 406:302–321.
- He, Kaiming, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2017a. "Mask R-CNN." In *IEEE International Conference on Computer Vision (ICCV)*, 2961–2969. IEEE.
- He, Kaiming, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2017b. "Mask R-CNN." In *Proceedings of the IEEE International Conference on Computer Vision*, 2961–2969.
- Huang, M., and C. Lee. 2024. "Cow face keypoint detection and head pose recognition using YOLOv7-Pose." *Journal of Animal Science and Technology* 66:45–59.
- Huelsmann, Axel, Ingo Buechner, and Ulrich Van Gemmeren. 2007. "Machine Vision: Laser projections assist machine-vision applications." *Laser Focus World*.
- Isensee, Fabian, Paul Jaeger, Simon Kohl, Jens Petersen, and Klaus Maier-Hein. 2021. "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation." *Nature Methods* 18 (2): 203–211.
- Jiang, Le, Caleb Lee, Divyang Teotia, and Sarah Ostadabbas. 2022. "Animal pose estimation: A closer look at the state-of-the-art, existing gaps and opportunities." *Computer Vision and Image Understanding* 222:103483.
- Jones, R., and S. Patel. 2019. "Sheep behavior and posture estimation using classical computer vision techniques." *Journal of Agricultural Engineering Research* 67:88–97.
- Kashiha, M., C. Bahr, S. Ott, C. P. H. Moons, T. A. Niewold, F. O. Ödberg, and D. Berckmans. 2014. "Automatic monitoring of pig locomotion using image analysis." *Computers and Electronics in Agriculture* 104:105–114.
- Kashiha, M., C. Bahr, and R. Wiegand. 2014. "Automatic animal detection in farm environments using classical methods." *Biosystems Engineering* 119:123–132. <https://doi.org/10.1016/j.biosystemseng.2013.11.004>.
- Kirillov, A., K. He, and R. Girshick. 2019. "Panoptic segmentation." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9404–9413. <https://doi.org/10.1109/CVPR.2019.00963>.
- Kirillov, A., E. Mintun, and N. Ravi. 2023. "Segment Anything Model (SAM) and its applications." *arXiv preprint arXiv:2304.02643*.
- Kirillov, Alexander, Kaiming He, Ross Girshick, Carsten Rother, and Piotr Dollár. 2019. "Panoptic Segmentation." *IEEE Conference on Computer Vision and Pattern Recognition*.
- Kirillov, Alexander, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, et al. 2023. "Segment Anything." *arXiv preprint arXiv:2304.02643*.
- Kumar, Abhishek, Ankit Singh, and Rakesh Sharma. 2020. "Semantic Segmentation of Iris using U-Net in Deep Learning." *International Journal of Recent Technology and Engineering* 9 (1): 2614–2618.

- Kuznetsova, Alina, et al. 2020. "The Open Images Dataset V4." *IJCV*.
- Kuzuhara, Yukako, Masami Osaka, Kazuo Senda, and Izumi Sato. 2015. "A preliminary study for predicting body weight and milk properties in lactating Holstein cows using a three-dimensional camera system." This study explores the use of 3D imaging for estimating body weight and related measurements in dairy cows. *Computers and Electronics in Agriculture* 117:186–190. <https://doi.org/10.1016/j.compag.2015.08.009>. <https://www.sciencedirect.com/science/article/pii/S0168169915002535>.
- Le Cozler, Yannick, C Allain, Anaïs Caillot, JM Delouard, L Delattre, T Luginbuhl, and Philippe Faverdin. 2019. "High-precision scanning system for complete 3D cow body shape imaging and analysis of morphological traits." *Computers and Electronics in Agriculture* 157:447–453.
- Li, Chen, and Gim Hee Lee. 2021a. "From synthetic to real: Unsupervised domain adaptation for animal pose estimation." In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1482–1491.
- Li, Chen, and Gim Hee Lee. 2021b. "From Synthetic to Real: Unsupervised Domain Adaptation for Animal Pose Estimation." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1482–1491. <https://doi.org/10.1109/CVPR46437.2021.00153>.
- Li, Chen, and Gim Hee Lee. 2023a. "Scarcenet: Animal pose estimation with scarce annotations." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17174–17183.
- Li, Chen, and Gim Hee Lee. 2023b. "ScarceNet: Animal Pose Estimation with Scarce Annotations." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. ArXiv:2303.15023.
- Lin, Tsung-Yi, et al. 2014a. "Microsoft COCO: Common Objects in Context." In *ECCV*.
- Lin, Tsung-Yi, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. 2017. "Feature Pyramid Networks for Object Detection." In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2117–2125. IEEE.
- Lin, Tsung-Yi, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. 2014b. "Microsoft COCO: Common Objects in Context." In *European Conference on Computer Vision*, 740–755. Springer.
- Liu, Shu, Lu Qi, Haifang Qin, Jianping Shi, and Jiaya Jia. 2018. "Path Aggregation Network for Instance Segmentation." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 8759–8768.
- Liu, Y., C. Wang, Q. Fu, and R. Kou. 2025. "Intelligent deep learning and keypoint tracking-based detection of lameness in dairy cows." *Veterinary Sciences* 12:1199.
- Liu, Y., and H. Wang. 2018. "Farm environment effects on vision-based livestock monitoring." *Biosystems Engineering* 167:45–57.
- Malek, Kaveh, Mahsa Sanei, Ali Mohammadkhorasani, and Fernando Moreu. 2025. "Increasing human immersion with image analysis using automatic region selection." *Expert Systems with Applications* 284:127938.

- Mathis, A., P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis, and M. Bethge. 2018. "DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning." *Nature Neuroscience* 21 (9): 1281–1289.
- Minaee, Shervin, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. 2021. "Image Segmentation Using Deep Learning: A Survey." *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Nye, H., and D. Turner. 2020. "Instance segmentation for livestock in multi-animal scenes." *Sensors* 20:4567. <https://doi.org/10.3390/s20164567>.
- Nye, J., P. Zingaretti, and M. Pérez-Enciso. 2020. "Review of computer vision techniques for cattle identification, tracking, and behaviour analysis." *Computers and Electronics in Agriculture* 173:105386.
- Nye, Jessica, Laura M Zingaretti, and Miguel Pérez-Enciso. 2020. "Estimating conformational traits in dairy cattle with DeepAPS: a two-step deep learning automated phenotyping and segmentation approach." *Frontiers in Genetics* 11:513.
- Oliveira, P., and R. Silva. 2017. "Tracking livestock movement using background subtraction and contour analysis." *Computers and Electronics in Agriculture* 140:12–22.
- Parivendan, S., and K. Rao. 2025. "Beyond single-animal detection: Multi-livestock pose estimation with deep learning." *Computers and Electronics in Agriculture* 210:75–90.
- PDR2020. 2020. *Grupo Operacional Preservar a Qualidade na Carne Arouquesa (PDR2020–101–031094)*. Medida 1. "Inovação" do Programa de Desenvolvimento Rural do Continente.
- Peng, Yifan, Haoyu Liu, and Wei Zhang. 2024. "Accurate cattle body keypoint detection for automated body measurement." *Animals* 14 (17): 2453. <https://doi.org/10.3390/ani14172453>.
- Pereira, Américo, Pedro Carvalho, Gil Coelho, and Luís Côrte-Real. 2019. "Efficient CIEDE2000-based color similarity decision for computer vision." *IEEE Transactions on Circuits and Systems for Video Technology* 30 (7): 2141–2154.
- Perneel, C., and J. Smith. 2025. "Consistent multi-animal tracking for livestock behavior analysis." *Journal of Agricultural Engineering Research* 115:45–60.
- Pezzuolo, Andrea, Luigi Sartori, and Francesco Marinello. 2018. "Exploiting low-cost depth cameras for body measurement in the livestock sector." This study evaluates low-cost depth camera approaches for livestock body measurement. *Computers and Electronics in Agriculture* 155:417–426. <https://doi.org/10.1016/j.compag.2018.11.030>. <https://www.sciencedirect.com/science/article/pii/S0168169918302850>.
- Porto, S., and J. Smith. 2013. "Classical image processing techniques for livestock segmentation." *Computers and Electronics in Agriculture* 95:50–60. <https://doi.org/10.1016/j.compag.2013.03.005>.
- Psota, P., and M. Kriška. 2020. "3D segmentation of livestock from multi-view images." *Computers and Electronics in Agriculture* 170:105233. <https://doi.org/10.1016/j.compag.2020.105233>.
- Quaresma, Miguel, Daniel Bacellar, Belén Leiva, and Severiano R Silva. 2019. "Estimation of live weight by body measurements in the Miranda donkey breed." *Journal of Equine Veterinary Science* 79:30–34.

- Rhodin, Helge, et al. 2018. "Vision-Based 3D Pose Estimation with a Camera Array." In *Proceedings of the European Conference on Computer Vision (ECCV)*.
- Rhodin, Helge, Mathieu Salzmann, and Pascal Fua. 2018. "Learning Monocular 3D Human Pose Estimation from Multi-View Images." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. 2015. "U-Net: Convolutional Networks for Biomedical Image Segmentation." In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 234–241. Springer.
- Ruchay, A., V. Kober, K. Dorofeev, and S. Kolpakov. 2021. "Precision livestock farming technologies for cattle monitoring and management: A review." *Biosystems Engineering* 209:162–180.
- Ruchay, Alexey, Konstantin Dorofeev, Vsevolod Kalschikov, Vladimir Kolpakov, Kinispay Dzhulamanov, and Hao Guo. 2021. "Live weight prediction of cattle using deep image regression." In *2021 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroA-griFor)*, 32–36. IEEE.
- Salau, Jennifer, and Joachim Krieter. 2020. "Instance segmentation with Mask R-CNN applied to loose-housed dairy cows in a multi-camera setting." *Animals* 10 (12): 2402.
- Salau, Joerg, and Joachim Krieter. 2020. "Automatic measurement of dairy cow body traits using computer vision." *Computers and Electronics in Agriculture* 173:105394.
- Sasaki, K., and M. Tanaka. 2025. "Mask-based individual animal segmentation for pose estimation." *Computers and Electronics in Agriculture* 210:108456. <https://doi.org/10.1016/j.compag.2025.108456>.
- Sasaki, Tohru, Toshitaka Sakai, et al. 2020. "Position and Posture Measurements Using Laser Projection Markers for Infrastructure Inspection." *Electronics* 9 (5): 807. <https://doi.org/10.3390/electronics9050807>.
- Shooter, Moira, Charles Malleson, and Adrian Hilton. 2021. "Sydog: A synthetic dog dataset for improved 2D pose estimation." *arXiv preprint arXiv:2108.00249*.
- Smith, J., and L. Brown. 2020. "Early vision-based detection of posture and locomotion in dairy cows using edge and contour features." *Computers and Electronics in Agriculture* 175:105–115.
- Smith, J., and R. Patel. 2025. "Ensemble learning for robust lameness detection using multi-keypoint motion features." *Computers and Electronics in Agriculture* 212:85–102.
- Sun, Haitao, Shuai Yang, Lijuan Chen, Pingyan Liao, Xiangping Liu, Ying Liu, and Ning Wang. 2023. "Brain tumor image segmentation based on improved FPN." *BMC Medical Imaging* 23 (1): 172.
- Tan, Mingxing, Ruoming Pang, and Quoc V. Le. 2020. "EfficientDet: Scalable and Efficient Object Detection." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10781–10790.
- Tangorra, F., and F. Rossi. 2013. "Shape-based methods for livestock silhouette segmentation." *Pattern Recognition Letters* 34:987–995. <https://doi.org/10.1016/j.patrec.2013.02.009>.

- Viazzi, Stefano, Gulnara Ismayilova, Mateusz Oczak, Lars T. Sonoda, Matthias Fels, Christoph Bahr, and Daniel Berckmans. 2014. "Automated detection of dairy cow posture using image analysis." *Computers and Electronics in Agriculture* 100:139–147. <https://doi.org/10.1016/j.compag.2013.11.005>.
- Wang, X., and Y. Zhao. 2025. "Generalization Challenges in Animal Pose Estimation." Forthcoming, *Computer Vision and Image Understanding*.
- Wang, Y., and L. Zhao. 2026. "Performance limitations of deep learning-based livestock keypoint detection in complex farm environments." *Biosystems Engineering* 235:22–36.
- Wang, Z., et al. 2024. "SN-FPN: Self-attention Nested Feature Pyramid Network for Image Segmentation." *IEEE Access*.
- Wei, L., and H. Zhang. 2023. "Spatio-temporal models for livestock locomotion analysis using keypoint trajectories." *Biosystems Engineering* 230:89–102.
- Wei, L., and H. Zhao. 2023. "CNN-based semantic segmentation for livestock images." *Computers and Electronics in Agriculture* 210:107921. <https://doi.org/10.1016/j.compag.2023.107921>.
- Yakheifer, P., and R. Singh. 2025. "Transformer-based livestock monitoring under complex farm conditions." *Computers and Electronics in Agriculture* 211:145–160.
- Yang, Yuxiang, Junjie Yang, Yufei Xu, Jing Zhang, Long Lan, and Dacheng Tao. 2022. "APT-36K: A large-scale benchmark for animal pose estimation and tracking." *Advances in Neural Information Processing Systems* 35:17301–17313.
- Yu, F., and X. Li. 2023. "Deep learning approaches to single-animal segmentation." *Artificial Intelligence in Agriculture* 14:101–113. <https://doi.org/10.1016/j.aiia.2023.02.005>.
- Yu, Hang, Yufei Xu, Jing Zhang, Wei Zhao, Ziyu Guan, and Dacheng Tao. 2021a. "AP-10K: A benchmark for animal pose estimation in the wild." *arXiv preprint arXiv:2108.12617*.
- Yu, Hang, Yufei Xu, Jing Zhang, Wei Zhao, Ziyu Guan, and Dacheng Tao. 2021b. "AP-10K: A Benchmark for Animal Pose Estimation in the Wild." *ArXiv:2108.12617, NeurIPS Datasets and Benchmarks Track*.
- Yu, Hang, Yufei Xu, Jing Zhang, Wei Zhao, Ziyu Guan, and Dacheng Tao. 2021c. "AP-10K: A Benchmark for Animal Pose Estimation in the Wild." Often cited for supervised performance evaluations, *NeurIPS Datasets and Benchmarks Track*.
- Yu, Hongshan, Zhengeng Yang, Lei Tan, Yaonan Wang, Wei Sun, Mingui Sun, and Yandong Tang. 2018. "Methods and Datasets on Semantic Segmentation: A Review." *Neurocomputing* 304:82–103.
- Yu, Y., C. Wang, Q. Fu, R. Kou, F. Huang, B. Yang, T. Yang, and M. Gao. 2023. "Techniques and Challenges of Image Segmentation: A Review." *Electronics* 12 (5): 1199.
- Zhang, Chaoning, Joseph Cho, Fachrina Dewi Puspitasari, Sheng Zheng, Chenghao Li, Yu Qiao, Taegoo Kang, Xinru Shan, Chenshuang Zhang, Caiyan Qin, et al. 2023. "A survey on segment anything model (SAM): Vision foundation model meets prompt engineering." *arXiv preprint arXiv:2306.06211*.
- Zhang, Haifeng, Yinan Miao, and Bo Yan. 2025. "On the Measurement of Laser Lines in 3D Space with Uncertainty Estimation." *Sensors* 25 (2): 298. <https://doi.org/10.3390/s25020298>.

- Zhang, Rui, Chen Li, and Jie Wang. 2023. "Vision-based facial keypoint detection for cattle behavior analysis." *Sensors* 23 (4): 2157. <https://doi.org/10.3390/s23042157>.
- Zhao, H., et al. 2026. "MambaFPN: Integrating State Space Models with Feature Pyramid Networks." *Neural Networks*.
- Zhao, K, JM Bewley, Dongjian He, and X Jin. 2018. "Automatic lameness detection in dairy cattle based on leg swing analysis with an image processing technique." *Computers and Electronics in Agriculture* 148:226–236.
- Zhou, T., W. Wang, F. Zhang, B. Chang, W. Wenguan, Y. Yuan, E. Konukoglu, and D. Cremers. 2024. "Image Segmentation in Foundation Model Era: A Survey." *arXiv preprint*.
- Zhou, Zongwei, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. 2020. "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation." *IEEE Transactions on Medical Imaging* 39 (6): 1856–1867.
- Zuffi, Sergio, et al. 2019. "Capturing Animals with a 3D Deformable Shape Model." In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.
- Zuffi, Silvia, Angjoo Kanazawa, David Jacobs, and Michael J. Black. 2019. "3D Menagerie: Modeling the 3D Shape and Pose of Animals." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.