# Final Report

## Computer Vision for Individual Animal Identification


Project code: V.DIG.0028

Prepared by: Georgia Mason, Darcy Joseph, Roland Croft, Brian Du
Swordfish Computing



Date published: 25th May 2026

PUBLISHED BY
Meat & Livestock Australia Limited
PO Box 1961
NORTH SYDNEY NSW 2059

Meat & Livestock Australia acknowledges the matching funds provided by the Australian Government to support the research and development detailed in this publication.



**This publication is published by Meat & Livestock Australia Limited ABN 39 081 678 364 (MLA). Care is taken to ensure the accuracy of the information contained in this publication. However MLA cannot accept responsibility for the accuracy or completeness of the information or opinions contained in the publication. You should make your own enquiries before making decisions concerning your interests. Reproduction in whole or in part of this publication is prohibited without prior written consent of MLA.**

# Abstract

Accurate cattle identification is essential for traceability in meat supply chains. Vision-based methods offer a non-intrusive alternative to RFID and can be combined with other computer vision tasks. This work investigates a fully compute -vision approach that identifies individual cattle using a multi-feature ensemble identification system. It matches each animal to the closest entry in a database of known identities.

A prototype ensemble was built and tested on two real-world datasets collected from research and development programs. On the controlled *Cattle Weight V.TEC.1731* dataset the recogniser achieved 100 % top-1 accuracy, demonstrating perfect identification under ideal conditions. On the more challenging *P.PSH.1424* dataset, designed to emulate practical farm environments, the prototype reached 62.5 % top-1 accuracy, confirming the feasibility of the approach while highlighting areas for improvement.

The project delivered a functional, expandable prototype and a comprehensive qualitative analysis of its strengths and limitations. Based on this analysis, a roadmap toward operational readiness is outlined, targeting a future deployment accuracy of 90 % through the integration of additional features and model refinements.

# Executive summary

## Background

The focus of this research project was to determine whether a fully computer vision-based system could accurately identify individual cattle with sufficient reliability to supplement traditional identification technologies such as RFID tags. A software-based solution of this kind offers a non-intrusive approach to cattle identification that can be automated without manual handling or data collection, while also integrating readily with other vision-based intelligence and supply chain tracking systems.

The project was primarily targeted toward the livestock and meat supply chain industries, including producers, researchers, and technology developers, due to the increasing demand for reliable traceability and automated livestock monitoring technologies.

Given the favourable results achieved during this research, the outcomes of the project are intended to support the development of scalable, non-intrusive computer vision-based cattle identification systems that are sufficiently robust, accurate, and efficient for operational deployment, while also supporting integration with broader agricultural computer vision applications.

## Objectives

The following objectives were achieved at the conclusion of this project, through the development, evaluation and delivery of a functional prototype recognition system:

- Review, analyse, and curate existing cattle image and video datasets to assess data quality and available visual features, while establishing a benchmark validation environment and evaluation protocol for recognition testing.
- Train and validate prototype computer vision recognition models capable of generating discriminative cattle identity embeddings using multiple visual cues, to assess the feasibility of AI-based cattle identification under both controlled and practical conditions.
- Design and implement a modular ensemble-based recognition system that combines multiple visual recognition features, to improve identification accuracy and robustness.
- Conduct a comprehensive performance analysis of the overall recognition system to evaluate operational feasibility, scalability, computational requirements, and deployment considerations.

## Methodology

The project followed an iterative machine learning workflow underpinned by agile and adaptive engineering practices, and was structured around the following key activities:

- Curated datasets were constructed from multi-camera video datasets through frame extraction, preprocessing, and automated segmentation, producing structured body and face feature datasets for model development.
- A multi-feature ensemble cattle identification system was designed to compare input images against a database of known cattle identities, with recognition models trained on the curated feature sets to perform identity matching.

- System performance was evaluated using top-1 identification accuracy to assess correct identification capability, alongside qualitative analysis to examine system strengths, limitations, and scalability for real-world deployment.

## Results/key findings

The prototype achieved 100% top-1 identification accuracy on the controlled Cattle Weight V.TEC.1731 dataset, demonstrating strong effectiveness under ideal imaging conditions. On the more complex P.PSH.1424 dataset, designed to reflect real-world farm environments, the system achieved 62.5% top-1 accuracy, confirming the feasibility of computer vision-based cattle identification while also highlighting the challenges introduced by operational variability. Importantly, the results indicate clear and achievable opportunities for further performance improvements through additional feature integration and model refinement with continued development effort.

## Benefits to industry

The research demonstrates the potential for non-intrusive, vision-based cattle identification systems to improve livestock traceability and automation within the meat supply chain. Such systems could reduce reliance on RFID technologies, lower handling stress for animals, and support integration with other automated computer vision applications such as behaviour monitoring and weight estimation. The expandable prototype also provides a foundation for future commercial and research development.

## Future research and recommendations

Future research should focus on improving robustness in real-world farm environments through the integration of additional visual features, more advanced data fusion strategies, and enhanced model refinement techniques. In addition, while this project was limited to frame-based image analysis, future work should extend to full video-based interpretation, incorporating temporal information through tracking and sequence-level modelling to better reflect operational deployment conditions. Further testing across varied lighting conditions, cattle breeds, and camera perspectives is also recommended to improve generalisability and operational reliability. Based on the findings of this project, achieving deployment-ready accuracy levels of approximately 90% appears feasible with continued development and optimisation.

# Table of contents

**Abstract** ..... 2

**Executive summary** ..... 3

**1. Project description** ..... 6

**2. Project objectives** ..... 7

**2.1 Milestone 1** ..... 7

**2.2 Milestone 2** ..... 7

**2.3 Milestone 3** ..... 8

**2.4 Milestone 4** ..... 8

**3. Methodology** ..... 8

**3.1 Agile research and development** ..... 8

**3.2 System overview** ..... 9

**3.3 Data preparation** ..... 10

**3.4 System validation** ..... 12

**3.5 Model training and identification** ..... 12

**4. Evaluation results** ..... 13

**4.1 Quantitative results** ..... 14

**4.2 Qualitative results** ..... 15

**4.3 Key findings** ..... 19

**5. Recommendations and Future Work** ..... 19

**5.1 System improvements** ..... 19

**5.2 Deployment considerations** ..... 20

**5.3 Future work** ..... 22

**6. Conclusion** ..... 23

**7. References** ..... 24

# 1. Project description

Individual animal identification of cattle and livestock is an important process that enables traceability and tracking of meat supply chains. The most common identification method enforced by Australia's National Livestock Identification System (NLIS), is to use electronic Radio Frequency Identification (RFID) ear tags that farmers scan with a handheld or automated panel readers to read identity (NLIS 2025). This method is labour intensive, invasive and requires proximity to cattle.

Recently, computer vision-based analysis of cattle has seen a significant rise in popularity due to advancements in robustness and accuracy of Deep Learning (DL) methods (Oliveira et al. 2021), this is evidenced by recent applications such as weight estimation (Gjergji et al. 2020), and cattle health diagnosis (Huang et al. 2019). The ability to estimate and track such characteristics through purely visual data provides significant utility to meat supply chains, by increasing available information and reducing labour. Currently, however, these metric estimators must be used alongside traditional cattle identification in order to be properly integrated into the supply chain.

Hence, this project aims to evaluate the feasibility of a purely vision-based solution for individual animal identification through development of an AI-enabled Cattle Identification System (ACIS) prototype. Such a system could easily integrate with other visual-intelligence systems (i.e. weight estimation models) and support existing hardware such as RFID tags and other sensors. Additionally, a digital system is highly flexible. New functionalities and improvements can be integrated as software updates, allowing farms to evolve without costly equipment changes.

Consequently, computer vision-based cattle identification has been a longstanding research challenge to help improve traceability and tracking of meat supply chains (Hossain et al. 2022, Meng et al. 2025). Existing methods exhibit substantial limitations when applied to real-world environments due to an over-reliance on single-feature approaches. These require high quality visual biometrics, such as facial imagery (Xu et al. 2022), or muzzle prints, which perform well in controlled environments where cattle are positioned in a crush to face a camera, but fail in real-world scenarios due to factors like noise, occlusion, and poor camera angles. There is a lack of investigation into computer vision-based cattle identification systems that rely on non-obtrusive visual data in open scenarios.

To address these challenges, this project set out to train an effective cattle recognition model through development of an ensemble-based cattle recognition system, that can integrate various features and visual indicators to make robust identifications. The use of a recognition system, in contrast to a classification system, enables the model to perform in an open-set setting to handle unknown identities and not require retraining when new identities are added. Additionally, the use of an ensemble-based recognition model enables the identification system to rely on a multitude of visual indicators to enable robust identification under diverse application setups. Such a design also enables easy deployment and enhancement, as future visual indicators can easily fit into the ensemble system.

# 2. Project objectives

The project objectives as outlined in the research agreement were as follows:

- Investigate and assess state-of-the-art AI and Computer Vision-based techniques for cattle identification.
- Evaluate the performance of selected AI techniques which may include facial, gait, and body-based recognition, determining the feasibility of a concept demonstrator system.
- Develop and validate a concept demonstrator showcasing early system capability, aiming for accuracy of over 80%.
- Provide recommendations for maturing and scaling the technology toward operational deployment, including infrastructure, data exchange interface recommendations.

The project's methodology was split into four key milestones, with the following goals and research aims:

## 2.1 Milestone 1

***Review Existing Dataset.*** Work on deliverable 1 will commence with a review of the existing dataset and its associated labels. This review focuses on:

- Data quality – camera angles, degree of occlusion, resolution, animal movements.
- Label format - definition and annotation of tracklet (a temporally contiguous series of frames in which an animal is detected and associated with the same identity, or provisional identity).
- Feature exploration – identification of available features, tracklet and class distribution.
- Multi-view synchronisation – verification that timestamps align across all video streams.

The outcome will be an analysis report that recommends whether further data collection or re-annotation is required and what are the available features for recognition.

***Dataset Curation.*** Building on the insights from the review, the Swordfish team will curate a dataset for training recognition models. The curation workflow includes:

- Extraction of the visual cues needed for the planned models (facial features, muzzle pattern, ear-tag markings, body coat patterns).
- Application of data augmentation techniques to improve generalisation.
- Integration of relevant open-source datasets when the existing data are insufficient, followed by a quality-control pass to ensure label consistency.

Curation will be iterative: after each training round we will evaluate performance on the benchmark gallery and, if the target accuracy is not met, we will augment the dataset accordingly.

***Validation Setup.*** To guarantee that the recognition system meets the intended use cases, a benchmark validation environment will be established:

- Benchmark test set – a representative collection of images/video clips that span the expected range of lighting, pose, occlusion and herd density.
- Evaluation protocol – definition of metrics (top-k accuracy) and a reproducible validation script that can be run automatically on any newly trained model.

This setup provides a repeatable basis for measuring system performance and for advising how different recognition subsystems can be orchestrated to achieve maximum accuracy.

## 2.2 Milestone 2

***Prototype Recognition Models.*** A prototype recognition system will be trained using metric-learning objectives such as triplet loss, contrastive loss or ArcFace, producing discriminative embeddings for

each animal. The model will be validated using the benchmark test set and evaluation protocol established in 4.2.1.3. The outcome of this task will indicate the feasibility of the proposed approach.

***Cattle Recognition System Feasibility Report.*** This milestone provides a focused assessment of the technical viability of AI-based cattle identification. Using the curated dataset and validation setup from Milestone 1, Swordfish will evaluate the performance and limitations of the proposed multi-factor recognition approach. The feasibility report will summarise model suitability, expected performance, operational requirements, and major technical risks, concluding with an overall viability assessment and recommendation on whether to proceed to prototype development.

## 2.3 Milestone 3

***Ensemble-based Recognition System Design.*** The design phase will start by identifying vision-based recognition technologies that best suit the curated dataset ranging from facial feature, muzzle pattern, emerging pattern, and body pattern recognition. A modular architecture will then be drafted that supports fusing of multiple model outputs; possible fusion strategies include score-level averaging, weighted voting or a learnable fusion layer.

***Training Recognition Models.*** Building on the prototype recognition system, this task explores different recognition models utilising other visual cues as outlined in 4.2.1.2. These new models will be validated both independently and as part of the ensemble using the benchmark test set and evaluation protocol established in 4.2.1.3. The deliverables for this sub-task are the trained recognition pipelines, training logs and performance reports.

## 2.4 Milestone 4

***Analysis Overall Recognition System Performance.*** The objective is to deliver a comprehensive assessment of the system's performance and identify its limitations to inform:

- Computation requirement
- Camera configuration
- Scalability & deployment environment

# 3. Methodology

This section outlines the methodology used to develop a prototype cattle identification system. The work follows a standard machine learning workflow, progressing from data preparation through to model development and evaluation. Initial data preparation enabled the construction of an ensemble-based recognition model, with subsequent stages refining performance through iterative experimentation.

The process was inherently iterative, with insights from feature engineering, training, and evaluation informing ongoing improvements and the development of a concept demonstrator. The following sections describe the development practices, system design, data preparation, modelling approach, and evaluation setup in more detail.

## 3.1 Agile research and development

This project was undertaken using fit-for-purpose engineering practices, adapted to the constraints of the environment. This includes:

- Agile and iterative engineering practices to enable flexible development cycles, regular planning and review, continuous stakeholder engagement, and incremental demonstrations of progress aligned with evolving project priorities.
- Configuration and task management through codebase version control, supporting task prioritisation, estimation, tracking, and scope management.

Knowledge management was conducted using a collaboration platform for information sharing and client communications.

## 3.2 System overview

This project developed a prototype concept technology demonstrator for computer vision–based cattle identification. An overview of the identification system pipeline is illustrated in **Figure 1**.

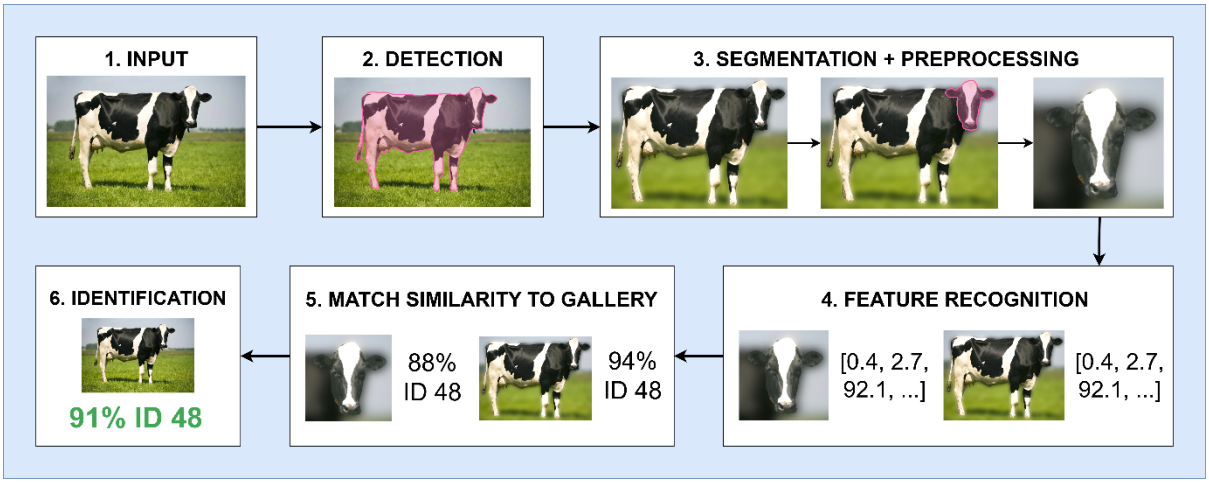



**Figure 1 - Diagram of proposed cattle identification system**

Rather than treating identification as a fixed classification problem, the system adopted a recognition-based approach. It learns to represent each animal as a distinctive 'feature embedding' signature, allowing new images to be matched based on similarity. An embedding is a numerical vector representing image features in a low dimensional space. This includes information such as shape, colour and texture. Identification of individual cattle is performed by comparing input images against a gallery of known registered cattle and selecting the closest match, as illustrated in **Figure 2**. This design enabled new animals to be added with minimal retraining, as only the gallery of known identities needed to be updated. Similarly, updating gallery images allowed the system to remain robust to temporal changes, such as animal growth, seasonal variation, and environmental conditions.

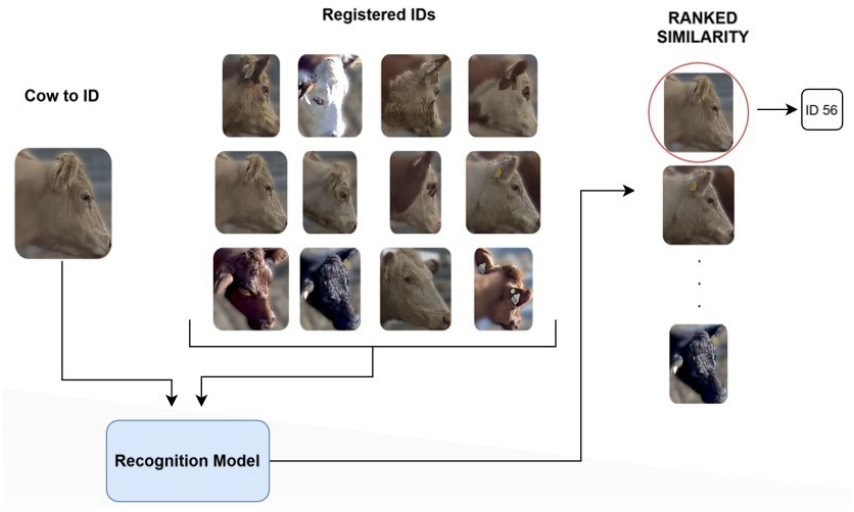



**Figure 2 – Cattle Recognition Model Pipeline**

To improve accuracy and robustness, the system integrated multiple visual features through an ensemble system design. Each model captured complementary aspects of the animal's appearance, and their outputs were combined to produce a final identification. This feature fusion approach increased reliability by leveraging multiple indicators, while the ensemble design allowed the system to remain effective even when input data was incomplete or of varying quality, by relying on the most informative representations available.

## 3.3 Data preparation

The project was conducted using a dataset provided by Meat & Livestock Australia and The University of New England through a related *P.PSH.1424* research and development program. The dataset comprised video recordings of steers moving through a race, captured from three distinct camera perspectives.

The *P.PSH.1424* dataset included side, front, and rear views of 105 individual cattle identities, organised across three camera streams: LOCO1(side), LOCO2(front/rear), and LOCO3 (front/rear). In total, the dataset contained 1,398 videos, with each identity represented by approximately 10–20 recordings of around 10 seconds in duration. Variability was introduced through differences in camera angle and collection day, providing a diverse set of imaging conditions for model development.

To prepare the data for training and evaluation, frames were extracted from each video at a rate of two frames per second. These frames constructed the image dataset used to train and evaluate the prototype identification system. The scope of this project was limited to image-based data for the training and evaluation of a computer vision model. Future work may extend this to include full video data, enabling a more comprehensive application. **Table 1** summarises the original video dataset and the resulting derived datasets.

**Table 1 P.PSH.1424 Dataset Overview**

| STATISTIC | | RAW DATASET | BODY DATASET | FACE DATASET |
|---|---|---|---|---|
| **NUMBER of IMAGES** | *LOCO1* | 9,292 | 6,098 | 4,055 |
| | *LOCO2* | 9,292 | 8,651 | 3,842 |
| | *LOCO3* | 9,333 | 8,380 | 3,123 |
| | *TOTAL* | 27,917 | 23,129 | 11,020 |
| **NUMBER of IDENTITIES** | | 105 | 105 | 105 |
| **IMAGES per IDENTITY** | *MEAN* | 265.9 | 219.9 | 104.6 |
| | *STD DEV* | 194.2 | 163.5 | 78.0 |
| | *MIN* | 27 | 20 | 8 |
| | *MAX* | 836 | 723 | 330 |

The extracted frames were passed through a preprocessing pipeline to produce two curated feature datasets: one focused on facial characteristics and the other on body-based features.

Effective feature representations are essential for distinguishing between individual cattle. In this work, two complementary visual features were selected: cattle face and body, allowing the impact of different inputs on identification performance to be evaluated. These features were chosen as they were the most straightforward to extract reliably and represented broadly generalisable cues for identity recognition. As such, they provided a strong basis for a proof-of-concept system, while remaining intentionally simple to support early-stage model development.

Facial regions were expected to provide fine-grained identifiers, such as ear shape, eye characteristics, and muzzle patterns. In contrast, body regions captured more holistic attributes, including overall silhouette, coat patterns, and muscle structure.

Feature extraction was performed using an automated segmentation pipeline to isolate the animal within each image. The process began by detecting the cow and extracting the corresponding region, which was then cropped, resized, and processed to reduce background influence. From this body segment, a corresponding facial region was subsequently extracted. The overall process is illustrated in **Figure 3**.

This dual-feature design supported an ensemble-based identification strategy, where complementary information from face and body derived representations could be combined to improve robustness and accuracy. While effective for a proof-of-concept system, future work would extend this approach to incorporate more sophisticated and diverse visual features to further enhance discriminative performance.




**Figure 3 – Image preprocessing and feature extraction pipeline (Note from *V.TEC.1731* data set)**

## 3.4 System validation

To train and evaluate the cattle identification system, the dataset was partitioned into training and testing subsets using a 70/30 split at the identity level, ensuring that cattle appearing in the test set were not seen during training. This allowed the model to be evaluated on unseen individuals, providing a realistic measure of generalisation performance.

The training set was used to train the recognition model, while the test set was further divided into two components: a gallery and a set of probe images. The gallery represented the known reference identities, while the probes represented query images to be identified. Model performance was then evaluated by determining whether each probe could be correctly matched to its corresponding identity within the gallery.

Both gallery and probe images were processed through the trained model to generate feature embeddings, which encode each image as a numerical representation of its visual characteristics. Identification was performed by comparing embeddings using cosine similarity, with each probe assigned the gallery identity with the highest similarity score. This matching process is illustrated in **Figure 4**.

To assess robustness under different operational conditions, three validation configurations were constructed using the *P.PSH.1424* dataset. The first configuration used all available camera views combined. The second and third configurations evaluated cross-view performance by using a single camera angle as the gallery and another as the probe set—specifically, LOCO2 (front view) versus LOCO3 (rear view), and vice versa. These setups were designed to simulate realistic deployment scenarios where camera perspectives may differ between enrolment and identification.

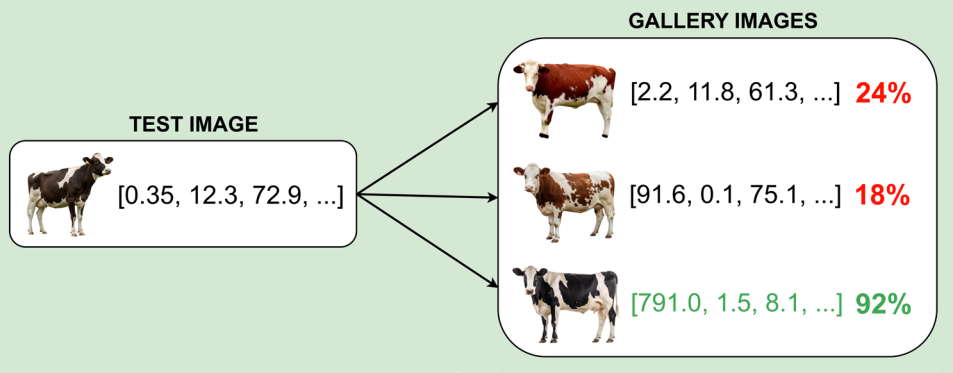



**Figure 4 – System validation setup.**

## 3.5 Model training and identification

The training dataset was used to develop a recognition model capable of generating feature embeddings that distinguish between individual cattle identities.

An initial approach was based on ArcFace (Deng et al. 2019), a widely used recognition framework originally developed for human face identification. This method applies metric learning with an additive angular margin loss to encourage the model to produce embeddings that maximise separation between different identities while maintaining consistency within the same identity.

However, this approach exhibited limitations in the context of the available dataset. Recognition models of this type are typically trained on very large-scale datasets to achieve strong generalisation

performance. In this case, the relatively limited number of training images constrained the model's ability to learn sufficiently robust representations. Additionally, large-scale data collection in cattle farming environments presents practical and economic challenges.

To address these limitations, a foundation model–based approach was adopted. A pretrained vision model was used as a feature extractor, leveraging representations learned from large and diverse image corpora. This reduced the reliance on extensive task-specific training data while still enabling the extraction of high-quality visual features.

As cattle identification requires fine-grained, instance-level discrimination, the pretrained model was further adapted using the available training data. An additional adapter module was trained on top of the base model to specialise it for cattle identification. The adapter was trained using an additive angular margin loss, encouraging improved separation between individual identities in the embedding space. This bespoke model architecture was termed *CowDINO*, reflecting its use of a DINO-based foundation model (Simeoni et al. 2025), adapted for cattle identification.

A body and face recognition model were independently trained, and then a multi-feature fusion approach was implemented to combine information from both representations. This design ensured that identification remained robust across varying conditions, particularly in cases where one feature was partially occluded or not clearly visible.

Several fusion strategies were considered. However, consistent with prior work in cattle identification, a late (score-level) fusion approach was selected for this implementation (Li et al., 2024). This approach aligned well with the training pipeline, in which separate models were developed for face and body features, and allowed each model to contribute independently to the final prediction.

Under this approach, face and body models were evaluated separately, producing similarity scores between each probe and gallery identity. These scores were then combined to produce a single fused score for each identity. The use of score-level fusion also enabled weighting to be applied to each feature, allowing the system to place greater emphasis on the more reliable representation under different conditions. The final identity prediction for each probe was determined based on the fused similarity scores, calculated as follows:

$$S(p,c) = \sum_{i=1}^{n} w_i \cdot \max_{m}\left(x_i(p,c,m)\right)$$

$$= w_{face} \cdot \max_{m}\left(x_{face}(p,c,m)\right) + w_{body} \cdot \max_{m}\left(x_{body}(p,c,m)\right)$$

Where ***p*** is the test probe, ***c*** is the cattle identity, ***n*** is the feature, and ***m*** is the set of test gallery images. $w_i$ represents the weight coefficient for the specific feature ***n*** (e.g. 0.5 for body and 0.5 for face) and $max(x_i)$ is the best similarity score for the particular **c** identity and ***p*** probe test.

# 4. Evaluation results

Both quantitative and qualitative analyses were conducted to evaluate the performance of the *CowDINO* recognition models trained on the *P.PSH.1424* dataset. Quantitative results were first assessed to determine overall performance within the target application scenario. This was followed by more detailed analysis to better understand model behaviour, including the structure of the learned feature representations and the nature of model predictions.

## 4.1 Quantitative results

Model performance was evaluated using similarity scores computed on the test set.

The primary metric was top-1 accuracy, defined as the proportion of test images for which the highest-ranked prediction matched the correct identity. This metric provided a direct measure of how often the system produced the correct identification.

The second metric was the Equal Error Rate (EER), defined as the point at which the False Acceptance Rate (FAR) and False Rejection Rate (FRR) were equal. This metric reflected the trade-off between incorrectly accepting an incorrect identity and incorrectly rejecting a correct one, with lower EER values indicating better overall discrimination.

The results from the *CowDINO* models are seen in **Table 2**. The best performing model was trained on all camera angles and applied feature fusion, to achieve a top-1 accuracy of 62.5%. The *CowDINO* model accuracy is significantly better than random guessing, which would be 3.13% for the 32-identity test set.

**Table 2: Metric results for *CowDINO* models on P.PSH.1424 dataset**

| Train + test sets | | | Metrics | |
|---|---|---|---|---|
| **Gallery** | **Probes** | **Feature** | **Top 1 (%)** | **EER** |
| **ALL** | **ALL** | *BODY* | 59.4% | 0.469 |
| | | *FACE* | 40.6% | 0.562 |
| | | *FUSION* | 62.5% | 0.375 |
| **LOCO2** | **LOCO3** | *BODY* | 23.5% | 0.656 |
| | | *FACE* | 24.5% | 0.648 |
| | | *FUSION* | 27.1% | 0.536 |
| **LOCO3** | **LOCO2** | *BODY* | 23.0% | 0.657 |
| | | *FACE* | 25.5% | 0.652 |
| | | *FUSION* | 26.9% | 0.518 |

It was observed that models trained and evaluated on single-location configurations (LOCO2 and LOCO3) performed substantially worse than the model evaluated across all camera views. This highlighted the challenges associated with transferring identification performance across differing imaging conditions and camera perspectives. In general, identification accuracy improved when the gallery reference images more closely resembled the probe images being evaluated.

To further investigate this behaviour, additional experiments were conducted using data provided through the *Cattle Weight V.TEC.1731* research effort. Unlike the *P.PSH.1424* dataset, this dataset contained only a single recording instance for each cattle identity, captured simultaneously across three camera angles. Under these conditions, the trained model achieved 100% top-1 accuracy using body features and 97.3% Top-1 accuracy using facial features.

This strong performance was attributed to the high similarity between the gallery reference images and the probe images, as both originated from the same recording sequence. The limited variation in environmental conditions, pose, and appearance allowed the model to identify highly consistent visual characteristics with minimal ambiguity. These results further emphasised the importance of representative gallery data and highlighted potential challenges in practical deployment scenarios where identification images may differ substantially from enrolment imagery.

Notably, feature fusion consistently improved performance across all three evaluation configurations, increasing top-1 accuracy by approximately 5-11%. Although the improvement was moderate, the results demonstrated the effectiveness of combining complementary visual features within an ensemble-based identification framework. Importantly, these gains were achieved using only two relatively simple feature types and a straightforward score-level fusion strategy, suggesting that more advanced fusion techniques and richer feature representations may yield further improvements.

To evaluate the benefit of the foundation model-based approach, an ablation study was conducted using the ArcFace architecture alone. In the best-performing configuration, this model achieved a top-1 accuracy of only 50%, representing a 21% reduction in performance relative to the equivalent *CowDINO* configuration. These results indicated that the foundation model-based approach produced substantially stronger visual representations and generalisation capability than the ArcFace-based approach alone.

## 4.2 Qualitative results

To further understand model behaviour and performance, qualitative analysis was conducted on the generated identification outputs. Manual visual inspection was performed on both the highest- and lowest-similarity matches within the test set.

Genuine identity matches generally achieved high similarity scores when strong visual alignment existed between the probe and gallery images. Images captured from highly similar viewing angles, and originating from the same recording instance, consistently produced the highest similarity scores, as seen in the **Table 3** examples. These observations further indicated that model performance was strongest when probe images closely resembled the gallery references, while performance degraded under variations in viewpoint and capture conditions.

For incorrect identifications, the selected “imposter” images frequently exhibited strong visual similarity to the input image despite belonging to different cattle identities. Similarities included coat colour, body shape, pose, viewing angle, and, in some cases, similar dirt patterns. The examples displayed in **Table 4** showcase the model’s bias to these factors. These results suggested that the model relied heavily on broader visual characteristics and would benefit from improved discrimination of more distinctive identity-specific features or the incorporation of more sophisticated feature representations.

Several poorly performing examples also highlighted the impact of visual obstruction. In the face-based model, lower-performing matches commonly occurred in side-view images where fencing partially obscured the animal’s face. Similar behaviour was observed in the body-based model, where key visual features were partially blocked by the crush structure. These observations indicated that occlusion and environmental obstruction represented significant challenges for reliable identification performance.

**Table 3: Example instances of high similarity matches**

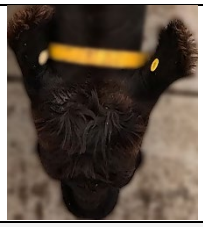
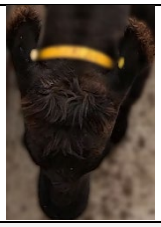
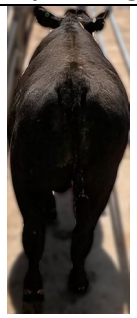
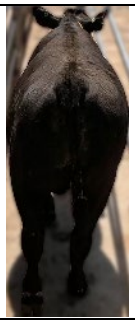
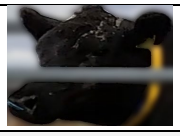
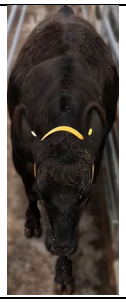
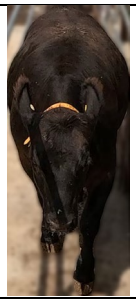
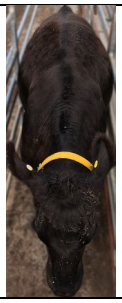
| POSITIVE MATCHES | | | |
|---|---|---|---|
| **Face model** | | | |
| **Identity** | **Similarity Score** | **Input image** | **Best match** |
| 028 | 0.657 |  |  |
| **Body model** | | | |
| **Identity** | **Similarity Score** | **Input image** | **Best match** |
| 084 | 0.999 |  |  |

**Table 4: Example instances of misclassifications**

| NEGATIVE MATCHES | | | | | | |
|---|---|---|---|---|---|---|
| **Face model** | | | | | | |
| **Input identity** | **Imposter identity** | **Imposter score** | **Best genuine score** | **Input image** | **Imposter** | **Best genuine** |
| 002 | 022 | 0.529 | 0.514 |  |  |  |
| **Body model** | | | | | | |
| **Input identity** | **Imposter identity** | **Imposter score** | **Best genuine score** | **Input image** | **Imposter** | **Best genuine** |
| 002 | 020 | 0.608 | 0.601 |  |  |  |

To further investigate the influence of visual characteristics on model behaviour, the clustering of latent feature embeddings was examined. As described in Section 3.5, the trained recognition model represented each identity as a feature embedding within a 512-dimensional latent space. By projecting these embeddings into a two-dimensional representation, a high-level visualisation of the model's ability to distinguish between cattle identities was obtained.

Two visualisations were generated using UMAP, a dimensionality reduction technique used to project high-dimensional embeddings into a two-dimensional space for visual analysis. The first visualisation was a colour-coded scatterplot of the resulting embeddings, where each colour represented a different cattle identity. This plot provided a high-level view of the model's ability to cluster identities in the learned feature space. The second visualisation was a connectivity graph, which illustrated relationships between points in the embedding space. In this representation, edges between points indicated local similarity structure, with connected clusters reflecting embeddings that were close in the higher-dimensional space. This helped to further interpret the structure of the learned representation, particularly in cases where clusters appeared visually separated but remained strongly connected through local neighbourhood relationships. Both visualisations were generated for the face-based *CowDINO* model.

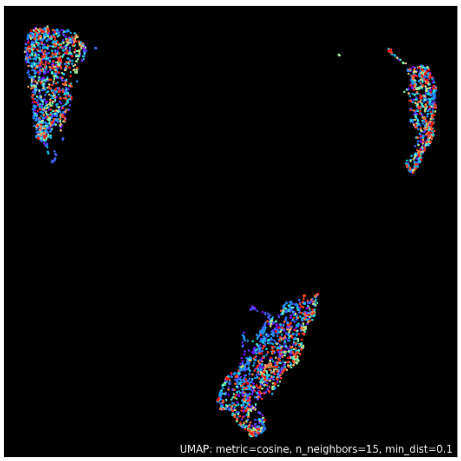



**Figure 5: Latent space scatterplot for the *CowDINO* face model**

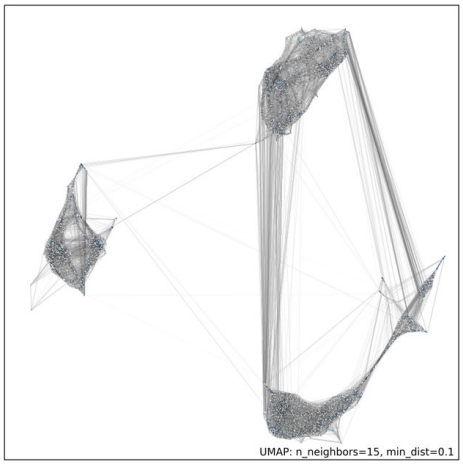



**Figure 6: Latent space connectivity graph for the *CowDINO* face model**

**Figure 5** shows the model split the data into 3 distinct clusters. These clusters align with the three distinct views of the cattle in the dataset, with one cluster consisting of only LOCO1 images, and other two were a mix of LOCO2 and LOCO3 images; split based on the direction the cow was facing. These results suggest the model's most distinguishable feature is the pose or direction the cattle are facing. However, the connectivity graph in **Figure 6** shows relationships between the cluster groups. This suggests that there are other features the model pays attention to; showcasing feasibility of this approach.

To further examine understanding of identity within a particular view, additional scatterplots were generated for only LOCO1 images.

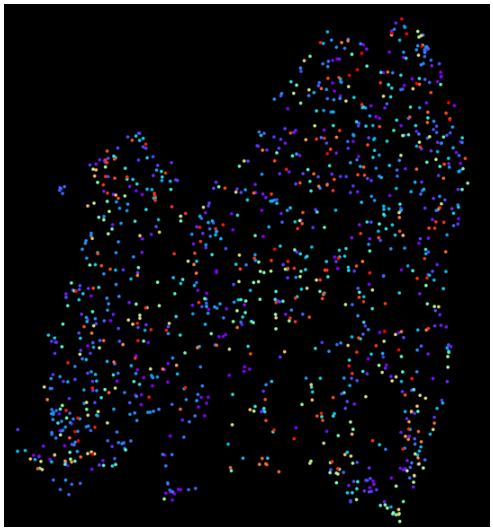


**Figure 7: Latent space scatterplot for *CowDINO* face model on LOCO1 data**

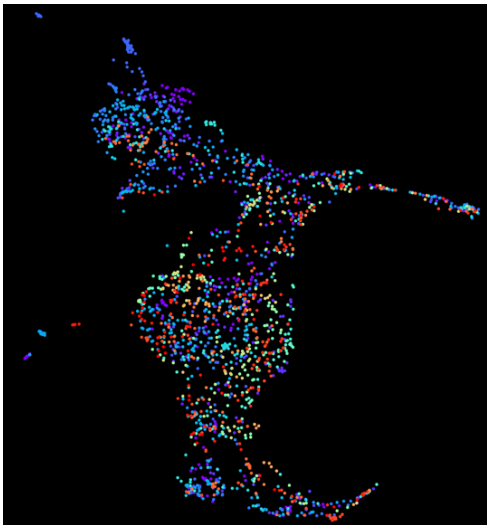


**Figure 8: Latent space scatterplot for *CowDINO* body model on LOCO1 data**

These visualisations further suggest that the model is strongly influenced by viewing angle, rather than finer-grained cattle-specific features, as identity clustering remains weak in the embedding space. The body-based model in **Figure 8** shows clearer structure, with some identities forming more distinct clusters, whereas the face-based model, **Figure 7**, exhibits limited separation between identities. This reinforces the benefit of an ensemble-based approach, particularly given that the camera setup was not optimised for high-resolution facial capture. As a result, body-based features provide more reliable signals and should be prioritised in the current configuration.

Overall, these results are consistent with the current multi-camera evaluation setup, which combines images from different viewing angles. Comparisons between differing viewpoints naturally produce lower similarity scores than comparisons between consistent views, which in turn limits overall identification performance. This suggests that future systems should either standardise comparisons to same-view configurations or incorporate mechanisms to account for viewpoint differences more explicitly.

### 4.3 Key findings

Several key findings were identified through quantitative and qualitative evaluation of the proposed cattle identification system. Overall, the results demonstrated the feasibility of a recognition-based approach for cattle identification, while also highlighting important challenges associated with real-world deployment conditions, feature representation, and environmental variability.

- The model achieved very strong performance under low-variance conditions, reaching up to 100% top-1 accuracy when gallery and probe images were highly similar. However, performance decreased substantially under more realistic conditions where viewing angle, appearance, and capture conditions varied, with accuracy reducing to approximately 62.5%.
- Camera angle was identified as a highly distinguishing visual characteristic within the learned feature space. As a result, identification performance was strongly influenced by viewpoint consistency, highlighting the importance of maintaining close alignment between gallery and probe image capture setups.
- Individual cattle exhibited high visual similarity, particularly in body colouration, shape, and pose. This limited the discriminative capability of simpler visual features and indicated the need for more sophisticated feature representations to improve identity separation.
- Facial features were found to be a comparatively weak predictor of identity within the current dataset and imaging setup. In particular, side-view imagery and partial occlusion from fencing significantly reduced the reliability of facial representations.
- The ensemble-based identification approach demonstrated clear value, with feature fusion consistently improving performance across all evaluation configurations. Combining complementary visual features increased robustness and reduced reliance on any single feature representation.
- The foundation model–based approach provided substantial performance benefits compared to the conventional ArcFace-based architecture. The pretrained model produced stronger visual representations and demonstrated improved generalisation capability under limited training data conditions.

## 5. Recommendations and Future Work

The evaluation results demonstrated the feasibility of a recognition-based ensemble identification system for cattle identification, while also highlighting several technical and operational challenges that would need to be addressed for practical deployment. Based on these findings, several recommendations were identified to improve system robustness, increase identification performance, and support transition from a proof-of-concept demonstrator toward a deployable operational capability. The following sections outline recommended system improvements, key deployment considerations, and potential directions for future development and application.

### 5.1 System improvements

While the prototype system achieved a top-1 accuracy of 62.5%, sufficiently demonstrating the feasibility of the proposed approach, the implementation intentionally prioritised simplicity to support proof-of-concept development. Commercial deployment within cattle farms and feedlots would require substantially higher reliability, likely exceeding 90% accuracy under varying

operational conditions. Achieving this level of performance would require improvements across data preparation, model architecture, feature representation, and fusion strategies.

The current preprocessing pipeline presents several opportunities for refinement. Although the segmentation approach produced accurate cattle regions, only simple prompting, thresholding, and background processing techniques were applied. Additional optimisation of these preprocessing parameters may improve segmentation quality and overall identification performance. Data augmentation was also used to increase training diversity. However, the applied techniques were tuned specifically to the current dataset and may not generalise effectively to broader operational conditions. Additional investigation into augmentation strategies better suited to foundation model architectures may therefore improve model robustness.

The underlying *CowDINO* architecture also remained relatively lightweight and minimally specialised toward cattle identification. While the adapter-based approach demonstrated strong performance improvements over conventional recognition architectures, further gains are expected through more advanced specialisation techniques, including:

- Increasing the complexity and representational capacity of the custom adapter module.
- Applying more advanced training strategies, such as staged training or full model fine-tuning.
- Expanding the scale and diversity of training data to improve generalisation capability.

The current system utilised only two feature types (face and body representations) which were combined using score-level feature fusion. Although this approach demonstrated measurable performance benefits, more sophisticated feature representations are expected to significantly improve identification capability. Potential future features include:

- Cattle gait analysis, using pose estimation to capture identity-specific movement patterns.
- Structural body measurements, such as hip height, body length, and body width, derived through pose estimation and depth sensing techniques.

As additional feature types are introduced, the fusion framework should also be extended. While the current score-level fusion method proved effective, future systems may benefit from learned fusion strategies that allow the model to dynamically determine the relative importance of each feature. Context-aware fusion approaches may be particularly valuable, enabling the system to adaptively prioritise the most reliable features when others are partially occluded or degraded.

## 5.2 Deployment considerations

The recommended deployment approach separates system use into two stages: a controlled gallery “onboarding” phase and a passive, real-world identification phase. During onboarding, each animal would be enrolled into the system by capturing high-quality gallery imagery under controlled conditions, allowing multiple clear views across relevant angles to be recorded. This capturing stage intentionally prioritises image quality and diversity, as it only requires to be performed once per animal and is critical for establishing a robust identity database. In contrast, live identification would operate in a non-intrusive manner using existing or minimally adapted stationed monitoring cameras positioned at common farm locations. These cameras would continuously observe animals in unconstrained conditions without requiring additional handling or disruption.

A key finding from this work is that identification performance is highly dependent on the consistency between gallery and live capture conditions; therefore, camera positioning, viewing

angles, and environmental setup during onboarding should closely match those used during operational deployment to ensure optimal system performance.

A production-level AI cattle identification system deployment will demand specific model and hardware requirements to facilitate this setup. Deployment will require computation hardware that can support continuous, multi video ingestion and inference with either an edge- or cloud-based compute setup. Further engineering work is required to evaluate the trade-off between edge and cloud-based compute setups for this specific application. The below hardware would be required at each capture station:

- High-Definition Electro-Optical (HD EO) Sensor:
  - *Used to record video sequences of cattle.*
- Mounting system for HD EO sensors:
  - *Fixed station at which HD EO sensor may be mounted.*
  - *Angle of sensor should be tuneable.*
- GPU-accelerated compute:
  - *Either edge- or cloud-based pending further engineering investigation and the specific compute requirements of models to be used.*
  - *Will also need to account for storage (e.g. hard drives) for video sequence storage.*
- Ruggedised containment:
  - *Waterproof, shock proof, lighting protected, etc.*
- Power supply

Based on the suite of models used in the cattle identification system prototype, the minimum compute and memory requirement for the three computer vision models are listed in Table 8. SAM3 is used for detecting and segmenting cattle during preprocessing, *CowDINO* is used for identification and YOLO26 is used for tracking cattle during a video sequence.

| COMPONENT | PARAMETERS | MIN RAM |
|---|---|---|
| **SAM3** | 848M | 16GB |
| **COWDINO** | 86B | 0.7GB |
| **YOLO26** | 2.4M - 55.7M | < 1 GB |

As *CowDINO* and YOLO26 both have multiple variants, the computational requirements of the identification pipeline may be adjusted by varying the size of the underlying vision model and the resolution of input imagery, although these reductions are generally associated with decreased identification performance. Within the current pipeline, the segmentation stage represented the primary computational bottleneck, and practical deployment would therefore require consideration of how frequently segmentation must be performed and whether the end-to-end pipeline is necessary at each video frame.

Fully real-time deployment, where segmentation and identification are performed continuously on high-frame-rate video streams, would require substantially greater computational capability. As such, practical deployment may require a balance between identification latency, processing frequency, hardware capability, and overall system cost. Further optimisation of the segmentation and inference pipeline would likely be necessary to enable efficient large-scale deployment in operational environments.

Alternatively, computational requirements may be reduced by performing segmentation intermittently rather than on every frame. Once a sufficiently confident identity prediction has been established for a tracked animal, segmentation and identification frequency could be reduced for the remainder of that track. This approach would significantly decrease processing demands while maintaining operational performance.

## 5.3 Future work

The current AI-based cattle identification system was designed with scalability and flexible deployment in mind. In a future implementation, the system would extend beyond image-based identification to incorporate full logic for establishing, tracking, and maintaining animal identities over time and across multiple camera views.

The prototype developed in this work treated identification primarily as an image-level feature matching problem. Although a video processing pipeline was implemented, individual frames were extracted and processed independently. In practical deployment, however, accurate identification requires reliable detection and tracking of individual cattle across consecutive frames, particularly in high-traffic environments such as feed troughs and water stations. A more robust future system would therefore integrate tracking to maintain consistent identity trajectories and select the highest-quality frames from each track for identification. A preliminary tracking pipeline was demonstrated using YOLOv26 for real-time detection and the BoT-SORT algorithm for multi-object tracking across frames. Further performance improvements could be achieved through fine-tuning of YOLOv26 and optimisation of tracking parameters using cattle-specific datasets.

While the current *CowDINO* model can be applied across different farms, its performance may be limited by dataset size and potential overfitting. For this reason, farm-specific retraining may be beneficial. Identity information collected during deployment could be used iteratively to update both the model and the gallery database. Successfully identified tracks would contribute additional high-quality images per identity, improving representation diversity and strengthening the underlying feature space over time. This would also help address appearance changes due to growth, seasonal variation, and environmental conditions by continuously updating gallery images with the most representative samples for each animal. In this way, the system would support a progressively improving identification pipeline that adapts to changing conditions in the field.

Finally, while the prototype system was primarily evaluated using top-1 accuracy, practical deployment may benefit from the use of an adjustable identification threshold to reduce misclassification risk and improve user trust. This threshold could be tuned based on operational requirements and validated using a dedicated calibration dataset derived from additional on-farm data.

# 6. Conclusion

The *V.DIG.0028 Computer Vision for Individual Animal Identification* project, a joint effort between Swordfish Computing and Meat & Livestock Australia, demonstrated that a computer-vision system can identify individual cattle in a controlled setting. Using a custom-curated dataset derived from the *P.PSH.1424* R&D program, the prototype achieved a top-1 accuracy of 62.4 %, confirming the technical feasibility of the approach.

The project delivered four key outcomes:

1. **Dataset curation pipeline** – a complete workflow for extracting frames from video, detecting and segmenting cattle, annotating individual animals for identification, and partitioning the data into training and validation sets.

2. **Validation framework** – a testing environment that reproduces field deployment conditions, allowing consistent performance measurement throughout development.

3. **Prototype ensemble system** – an ensemble of recognition models built around a novel architecture that couples a foundation model with additive angular margin loss, delivering the reported accuracy.

4. **Performance analysis and recommendations** – a thorough evaluation using metrics such as top-1 accuracy and equal error rate, visual inspection of positive and negative examples, and latent space visualisation with UMAP. The analysis produced concrete recommendations for advancing from prototype to commercial product.

The analysis also identified the hardware and infrastructure required for farm-level deployment. A modest edge-computing device paired with a high-resolution, low-latency camera, was found sufficient to run the ensemble in real time. Network requirements are minimal, as inference can be performed locally with periodic synchronization to a central server for model updates.

In summary, the prototype validates the technical viability of automated cattle identification, delivers a robust validation pipeline, and outlines the hardware and deployment strategies needed for commercial adoption. This foundation will enable future MLA research initiatives to supplement current identification methods with a scalable, non-invasive solution.

The project team would like to acknowledge the University of New England who assisted project delivery by providing access to the *P.PSH.1424* dataset.

# 7. References

Carion, N., Gustafson, L., Hu, Y.T., Debnath, S., Hu, R., Suris, D., Ryali, C., Alwala, K.V., Khedr, H., Huang, A. and Lei, J., 2025. Sam 3: Segment anything with concepts. *arXiv preprint arXiv:2511.16719*.

Deng, J., Guo, J., Xue, N. and Zafeiriou, S., 2019. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4690-4699).

Gjergji, M., de Moraes Weber, V., Silva, L.O.C., da Costa Gomes, R., De Araújo, T.L.A.C., Pistori, H. and Alvarez, M., 2020, July. Deep learning techniques for beef cattle body weight prediction. In *2020 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.

Hossain, M.E., Kabir, M.A., Zheng, L., Swain, D.L., McGrath, S. and Medway, J., 2022. A systematic review of machine learning techniques for cattle identification: Datasets, methods and future directions. *Artificial Intelligence in Agriculture*, *6*, pp.138-155.

Huang, G.B., Mattar, M., Berg, T. and Learned-Miller, E., 2008. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*.

Huang, X., Hu, Z., Wang, X., Yang, X., Zhang, J. and Shi, D., 2019. An improved single shot multibox detector method applied in body condition score for dairy cows. *Animals*, *9*(7), p.470.

Li, D., Li, B., Li, Q., Wang, Y., Yang, M. and Han, M., 2024. Cattle identification based on multiple feature decision layer fusion. *Scientific Reports*, *14*(1), p.26631.

Meng, H., Zhang, L., Yang, F., Hai, L., Wei, Y., Zhu, L. and Zhang, J., 2025. Livestock biometrics identification using computer vision approaches: A review. *Agriculture*, *15*(1), p.102.

National Livestock Identification System (NLIS). Integrity Systems. https://www.nlis.com.au/

Oliveira, D.A.B., Pereira, L.G.R., Bresolin, T., Ferreira, R.E.P. and Dorea, J.R.R., 2021. A review of deep learning algorithms for computer vision systems in livestock. *Livestock Science*, *253*, p.104700.

Siméoni, O., Vo, H.V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khalidov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M. and Massa, F., 2025. Dinov3. *arXiv preprint arXiv:2508.10104*.

Wang, Z.J., Turko, R., Shaikh, O., Park, H., Das, N., Hohman, F., Kahng, M. and Chau, D.H.P., 2020. CNN explainer: learning convolutional neural networks with interactive visualization. *IEEE Transactions on Visualization and Computer Graphics*, *27*(2), pp.1396-1406.

Xu, B., Wang, W., Guo, L., Chen, G., Li, Y., Cao, Z. and Wu, S., 2022. CattleFaceNet: A cattle face identification approach based on RetinaFace and ArcFace loss. *Computers and Electronics in Agriculture*, *193*, p.106675.