Exploring a Siamese Neural Network as a novel individual cow identification technique for daily monitoring free water intake

Matteo Di Vincenzo

Miel Hostens

Abstract

Water is an essential nutrient for a healthy and productive dairy herd, as most life processes require water. As cows produce more milk, their water requirement increases. Therefore, insight into the drinking behavior of dairy herds is important, but the data currently available is outdated. Furthermore, although the importance of water is globally recognized, the daily water intake of individual cows is currently not monitored, making it impossible to ensure that all cows meet their water needs. Efficient techniques to monitor the water intake of individual cows are highly needed. Computer vision can be used to complete almost all the tasks required for such monitoring job. The process requires detection of drinking behavior and identification of individual cows. The approach involves training a Siamese neural network with triplet loss to identify individual cows. This network generates feature vectors (embeddings) from images, ensuring that embeddings of the same cow are close while those of different cows are far apart.

Ultimately, the neural network can learn to distinguish coat patterns on cows, rather than directly learning identifiers that correspond to images. After training, the model can distinguish the spot patterns of cows well enough to need one or few images per cow to correctly classify all the cows in the herd. Therefore, this classification is called one-shot classification. Inference involves computing embeddings for new images and making predictions based on the most similar embeddings in the database. The model however had a fairly low accuracy when testing, and upon closer inspection, it seemed to be affected by the image's background. The proposal to tackle this problem has been addressed by using the instance segmentation model SAM by meta-AI along with GroundingDINO for zero-shot detection, in order to build a model that could correctly spot a cow in each image and subsequently segment it, leading to a clear image. Although the model now looks promising, further work is required to validate the model on the images without background. Once achieved, our model can work with pose estimation systems like DeepLabCut to monitor cow drinking behavior and assess individual water needs on the farm.

Layman’s Summary

A dairy cow's milk quantity and quality depend on the amount of water it consumes. In recent decades, there has been a significant increase in milk production per cow. As a result, the water requirements for each cow have risen. However, we currently lack a system to precisely monitor individual cow water intake on dairy farms. To address this issue, we aim to harness the power of artificial intelligence (AI). We plan to install cameras near the cow drinking troughs to observe their drinking behavior. These cameras will employ advanced algorithms like GroundingDINO to detect when a cow is present in an image and, to further determine if the cow is actively drinking, an algorithm called DeepLabCut.

To ensure accurate tracking of each cow's water consumption, we've developed a model called a "Siamese Neural Network"(SNN). This unique AI system is capable of two crucial tasks: distinguishing between different cows in images and recognizing the same cow across various images. By integrating these components, we can analyze video footage and precisely identify when a cow is drinking, and which specific cow is doing so. It's important to note that our Siamese Neural Network has been improved to remove background elements, leaving only the cow visible in the image, leading to more accurate results.

However, while our system shows promise, it still requires further training using a larger dataset. Once this is complete, we will be able to accurately measure the duration of time each cow spends drinking. This information, combined with data from water flow meters, will enable us to closely monitor the quantity of water each cow consumes. Ultimately, this approach will ensure that our cows receive the optimal amount of water necessary for high-quality milk production.

1. Introduction

1.1 Water Consumption in Dairy Cattle

Water is widely acknowledged as a pivotal factor in the health and productivity of dairy herds (Houpt, 1984). Within the context of dairy nutrition, water is often designated as the quintessential nutrient, essential for facilitating critical physiological processes. These processes encompass digestion, nutrient metabolism, thermoregulation, maintenance of fluid and ion equilibrium, and waste elimination. To underscore its significance, even a relatively modest 20% loss of body water can be fatal for dairy cattle. Insufficient access to water carries tangible consequences for dairy herd performance. Studies conducted by Little et al. in 1980 demonstrated a notable decrease in milk production when adult cattle were subject to water deprivation. Furthermore, Kertz, Reutzel, and Mahoney's research in 1984(Kertz et al., n.d.) highlighted that inadequate water intake in calves correlates with suboptimal weight gain. Notably, milk yield per cow per year has increased from 2580 to 9200 kg in the Netherlands between 1910 and 2022 (*How Much Milk Does a Cow Produce?*, 2023), from 6700 to 9300 kg in the United States and from 4900 to 6600 kg in Germany between 1990 and 2009 (Zehetmeier et al., 2012). This substantial upsurge in milk production per cow raises pertinent questions concerning the adequacy of water resources to sustain such elevated levels of production considering that milk is for roughly 87% made of water.

Beyond the immediate impact on productivity, ensuring a dependable water supply for dairy cattle is increasingly recognized as a matter of welfare and sustainability. Ongoing concerns associated with climate change (Henning Steinfeld, 2006) and water scarcity (Ridoutt et al., 2010; Cardot et al., 2008) have prompted a re-evaluation of the dairy industry's environmental footprint. Concurrently, attention to the health and well-being of dairy cows has grown (Barkema et al., 2015).

A conspicuous trend within the dairy sector is the substantial increase in milk production per cow in the last decade (Zehetmeier et al., 2012). This upsurge raises pertinent questions concerning the adequacy of water resources to sustain such elevated levels of production. Consequently, it is imperative to institute a rigorous system to monitor water consumption of dairy cows. This urgency is further underscored by the observed disconnect between extant data and the contemporary milk production per cow, necessitating updated and meticulous water intake monitoring to uphold the health and productivity of dairy herds (Houpt, 1984; Murphy, 1992.; Jensen & Vestergaard, 2021).

1.2 Automation of Free Water Intake Monitoring and Addressing the Identification Challenge

In recent decades, we've witnessed remarkable technological progress, particularly in the realm of automation. These advances have transformed the landscape of surveillance and detection across various sectors, with agriculture standing out as a prime beneficiary. Automation doesn't just simplify once laborious data collection tasks; it also empowers us to monitor vital parameters that were once challenging to assess manually. One such parameter is the water consumption of dairy cows, a critical aspect of their well-being (Cardot et al., 2008; Meyer et al., 2004).

While the significance of water consumption for animal welfare is widely recognized, there remain significant gaps in our ability to routinely monitor individual Free Water Intake (FWI) (Cardot et al., 2008). Historically, FWI measurements were confined to experimental settings and were not seamlessly integrated into daily monitoring practices. In experimental contexts, FWI has been quantified through automatic water flow registration (Melin et al., 2005) or by manually reading water meters associated with individual stalls(Huuskonen et al., 2011); (Liang et al., 2020).

Concurrently, an equally pivotal facet of effective monitoring is the precise identification of individual cows within a dairy herd. Traditionally, Radio-Frequency Identification (RFID) has been the go-to method for animal identification (Mendes et al., 2011; (Matthews et al., 2016). However, despite its widespread adoption, RFID has inherent limitations. Initially, it involves an invasive data gathering process, mandating the physical attachment of an RFID tag to each and every animal. Thereafter, every cow necessitates a distinct RFID tag, resulting in increasing expenses that correlate with the herd's size. Lastly, RFID relies on proximity, demanding the deployment of RFID scanners at multiple locations for data collection.

To comprehensively overcome these challenges and present an all-encompassing solution, we advocate for the pioneering implementation of computer vision technology. This non-intrusive method not only streamlines the identification of individual cows within a dairy herd but also carries the potential for seamlessly automating FWI monitoring. This holistic approach holds the promise of optimizing dairy herd management while ensuring consistent access to an adequate water supply for every dairy cow.

1.3 Introduction to the Application of Computer Vision

In the expansive landscape of scientific inquiry, computer vision emerges as an interdisciplinary domain, strategically poised to replicate the complexities of the human visual system. This endeavor empowers computers with the ability to perform tasks reminiscent of human capabilities (Huang, 1996). Within the tapestry of this multifaceted discipline lies a diverse array of subtasks, ranging from object detection to object recognition and pose estimation.

Our focus is on Grounding DINO, a groundbreaking system that redefines the possibilities of object detection and recognition (Liu et al., 2023). This innovative solution marries the robust capabilities of the Transformer-based detector DINO with grounded pre-training. The result is a system with the remarkable capacity to identify diverse objects, even when guided by human inputs such as category names or referring expressions. Grounding DINO's charm surpasses its competence as it stands as the only zero-shot detection system with an impressive 52.5 Average Precision (AP) score on the COCO detection zero-shot transfer benchmark. Moreover, its open-source architecture offers unparalleled versatility, facilitating seamless adaptation for training on specialized datasets (Liu et al., 2023).

In contrast, pose estimation, a critical facet of computer vision, finds its embodiment in DeepLabCut (DLC) (Mathis et al., 2018). This tool is complemented by a user-friendly graphical interface (Nath et al., 2019), greatly simplifying the labor-intensive tasks of data labeling and model training. These innovative methodologies provide the bedrock for the development of customized systems designed to monitor drinking behavior through the lens of computer vision. For example, DLC empowers users to precisely determine the spatial positions of various segments of a cow's anatomy. These spatial cues are instrumental in discerning whether a cow is actively engaged in the act of drinking. When a cow's mouth, for instance, approaches a water trough, it signifies a higher likelihood of the cow assuming a drinking posture compared to when the mouth is distanced from the trough.

However, the task of monitoring drinking behavior necessitates the accurate identification of cows during the act of drinking. To address this challenge comprehensively, we advocate for the utilization of Siamese Neural Networks (SNN), a promising solution within the realm of computer vision.

1.4 Siamese Neural Network (SNN)

A Siamese Neural Network (SNN) exemplifies the concept of similarity learning, distinguished by a two-stage process depicted in Figure 1. In the training stage, instead of directly associating labels with images, the SNN focuses on learning to distinguish between different objects without predicting their correct labels. This is akin to how a human can visually differentiate between, say, an elephant and a monkey without necessarily knowing their names. Only after becoming adept at distinguishing individual cows, does the model employ the corresponding labels in the inference stage to classify new images. During the training stage, the inclusion of more images per cow and a greater overall number of images accelerates the learning process. However, once the inference stage is reached, only one or a few images per cow are sufficient for accurate classification. In essence, the SNN becomes a one-shot classifier, requiring minimal examples to classify a cow correctly (Koch et al., 2015; Schroff et al., 2015).

The SNN operates by converting images into n-dimensional feature vectors, also known as embeddings, through convolution. In the training stage, it learns to make embeddings of the same cow more similar while making embeddings of different cows less similar. In the inference stage, classification involves comparing the embedding of a new cow image with the embeddings of cow images in a database, with the most similar embedding likely belonging to the same cow (Koch et al., 2015.; Schroff et al., 2015). One significant advantage of the SNN is its reduced need for retraining when introducing a new cow to the herd. Rather than learning specific spot patterns, the model learns to distinguish different spot patterns. Assuming the model generalizes effectively, only one image of a new cow is required to update the database without extensive retraining. However, effective generalization relies on training the SNN on a sufficiently large dataset that encompasses the full spectrum of possible variations in cows' spot patterns (Wang et al., 2014; Schroff et al., 2015). Consequently, an SNN trained on herd A may not seamlessly distinguish all cows in herd B, but with an increasing number of cows in herd A, its ability to differentiate cows in herd B should improve.

Triplet Loss

The SNN is often visually depicted as a network with multiple identical Convolutional Neural Networks (CNNs) stacked on top of each other. While this visualization aids in understanding, it's essential to note that there is only one CNN. Comparing the outputs of this CNN is achieved through a concept known as triplet loss. Triplet loss groups three embeddings into triplets: an anchor (x_a), a positive (x_p), and a negative (x_n). The anchor (x_a) and positive (x_p) are embeddings from the same cow, while the anchor (x_a) and negative (x_n) are from different cows. Triplet loss penalizes both large distances (D) between embeddings of the same cow, D (x_a, x_p), and small distances between embeddings of different cows, D (x_a, x_n). This triplet organization motivates the stacked visualization.

Offline Triplet Generation

Two main methods exist for implementing triplet loss: offline and online (Schroff et al., 2015). Offline triplet generation occurs before or between training epochs. Data is organized into triplets (x_a, x_p, x_n), split into batches, and then training commences. While this approach offers swift computation, its performance can deteriorate over time as the model optimizes weights to make D (x_a, x_p) small and D (x_a, x_n) large, gradually losing information (Schroff et al., 2015). This limitation can be mitigated by periodically regenerating triplets every few epochs (Schroff et al., 2015).

Online Triplet Generation

In contrast, online triplet generation occurs within mini-batches to identify the most informative triplets after converting them into embeddings. Each embedding can serve as an anchor (x_a), and the positive (x_p) with the largest D (x_a, x_p) and the negative (x_n) with the smallest D (x_a, x_n) within each batch have the most significant impact on training (Schroff et al., 2015). Although this approach consumes more time, it generally yields better performance gains (Schroff et al., 2015).

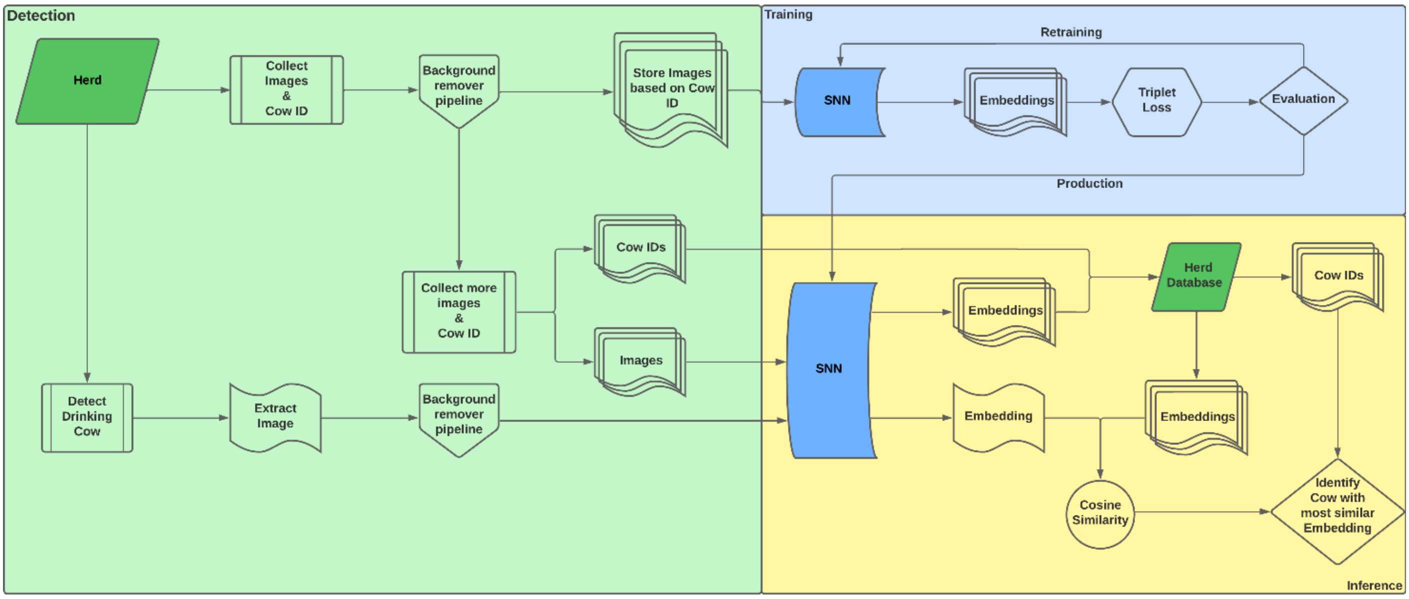

Fig. 1:

Proposed workflow of the drinking monitor

2. Materials and methods

2.1 Datasets

For our study, a surveillance system on a Dutch dairy farm was set up, equipped with six strategically positioned cameras, each focused on a different drinking trough. On October 10, 2022, 24 hours of continuous video footage were collected. This dataset serves as the foundation for our research, allowing us to investigate dairy cow behavior and water consumption patterns. In the following sections, we'll detail our analysis methods to gain insights into the datasets structure.

DONE dataset

On the same farm, we extended our surveillance efforts to encompass the automated milking robots. Four cameras were strategically installed, situated at the robot entrances and directed downward to capture the top view of each cow. This initiative resulted in the recording of a total of 221 cows during their milking sessions. From this extensive dataset, we manually extracted twenty images for each of the 170 cows, out of the 221, during a milking session. It's worth noting that the reason for the reduction of cows used was essential due to the time-consuming nature of data extraction. These images were sourced from three distinct milking robots, ensuring diversity within the dataset. In summary, our dataset, denoted as "DONE," comprises a collection of 3,400 images, featuring 170 distinct cows. To facilitate model training and validation, we divided this dataset into an 80% training subset and a 20% validation subset. Additionally, because all frames in the video recordings were equipped with timestamps and cam-labels, we applied blurring techniques to these in all images preventing the model from learning the timing or camera sources.

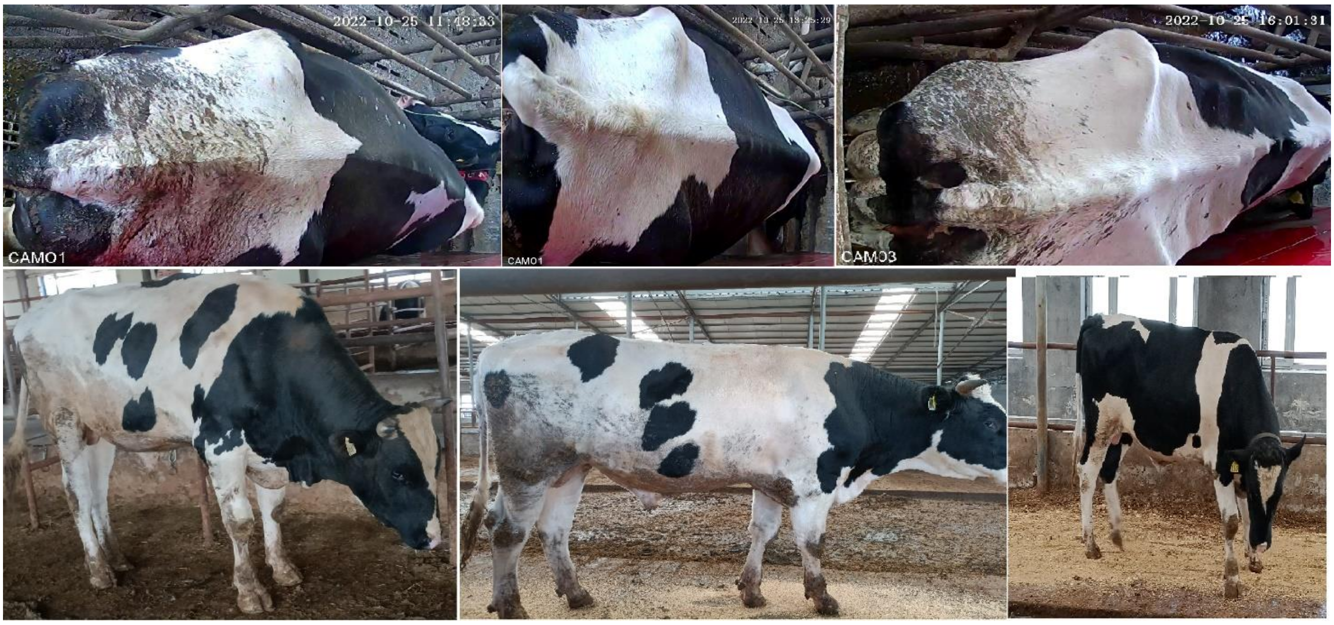

Fig. 2:

Comparison of images from the 2 datasets.

IC-64

A subset of a dataset published by (Li et al., 2021), comprises 1,040 high-quality images featuring 13 distinct cows. Each cow is represented by 80 images, ensuring diverse angles, lighting conditions, and backgrounds. To maintain equal image representation for our model, we utilized this subset, splitting it into an 80%/20% train-validation ratio. This selection was made to assess the potential impact of dataset quality on model performance while maintaining a balanced class distribution.

2.2 Background work and Model architecture

The model architecture comprises two main stages: feature extraction and feature mapping. In the first stage, multiple convolutional layers are employed to extract features from the input image. In the second stage, these extracted features are transformed into a 128-dimensional vector. The model, previously initiated by Daniël M. van Herwijnen, is derived from Google's InceptionResnet (Szegedy et al., 2016), with a notable modification. Unlike InceptionResnet, which utilizes pretrained weights, this adaptation initializes the weights randomly. Following the Inception network, the architecture incorporates two fully connected layers, each consisting of 1028 neurons, and a final fully connected layer that produces a 128-dimensional vector. To enhance the robustness of the model, the input data undergo augmentation through random adjustments in brightness, contrast, and translation. The translation layer introduces horizontal and vertical shifts to reduce the model's dependency on the precise positioning of the cow within the image.

Subsequently, the model underwent significant refinements to enhance its adaptability and adherence to best practices, aligning it more closely with the FAIR (Findable, Accessible, Interoperable, and Reusable) principles (Wilkinson et al., 2016) by refining the documentation. These refinements included the addition of comprehensive comments throughout the code, improving readability and making the implementation more transparent to future developers and collaborators. These efforts aimed to facilitate easier access, understanding, and reusability of the model. Although the core architecture remained intact, substantial modifications were introduced to improve its functionality.

Originally, the model relied on predetermined pathways and GPU resources for processing, applying a uniform augmentation process to all images. However, these limitations have been addressed and rectified in the following ways:

Dynamic Resource Allocation

The model's adaptability has been greatly enhanced by implementing dynamic resource allocation. It can now detect available GPUs or TPUs and gracefully fall back to CPU processing if none are found. This ensures that the model can efficiently utilize the computing resources at its disposal.

Flexible Directory Structure

The rigid directory structure has been revamped to operate within a more flexible environment. This change allows for greater versatility in handling various file structures and locations, making the algorithm more adaptable to different data setups.

Randomized Augmentation

The augmentation process has been redesigned to introduce randomness when applied to different images. Rather than a uniform augmentation approach, the algorithm now initiates diverse augmentation techniques for individual images. This enhances the model's versatility and usability across a wide range of datasets.

Improved Documentation

Recognizing the significance of comprehensive documentation, considerable efforts have been invested in enhancing the code's clarity and accessibility. This endeavor aims to make the codebase more transparent and user-friendly, aligning it better with FAIR principles (Wilkinson et al., 2016). By providing more comprehensive documentation, we seek to facilitate a deeper understanding of the code's functionalities and ensure a smoother user experience.

Enhanced Functionality

Several enhancements have been made to the codebase to improve efficiency and accommodate more flexible data handling. This includes transitioning from fixed quantities to percentage-based data handling and incorporating reminder logic for batch creation.

These extensive refinements collectively contribute to a more adaptable, user-friendly, and FAIR-compliant model architecture, poised to meet the evolving needs of scientific research and application.

2.3 Background removal pipeline

After the initial training, some improvement in the model's accuracy was observed. However, further investigation was deemed necessary, as suggested by Daniël M. van Herwijnen's hypothesis, to identify potential factors affecting performance. A critical consideration in this context was the presence of background elements in the training images. Upon closer examination, when the trained model was evaluated on the validation dataset, it exhibited some degree of accuracy, albeit marginally above that of a random classifier. However, a noteworthy anomaly surfaced when the same set of images, with backgrounds removed, was employed for evaluation. In this scenario, the model's performance suffered a substantial decline, confirming the hypothesis that the model primarily recognized the entire image rather than accurately detecting and recognizing the cow. This limitation became especially apparent due to the similarity in backgrounds across all images, which hindered the model's ability to effectively identify the cows.

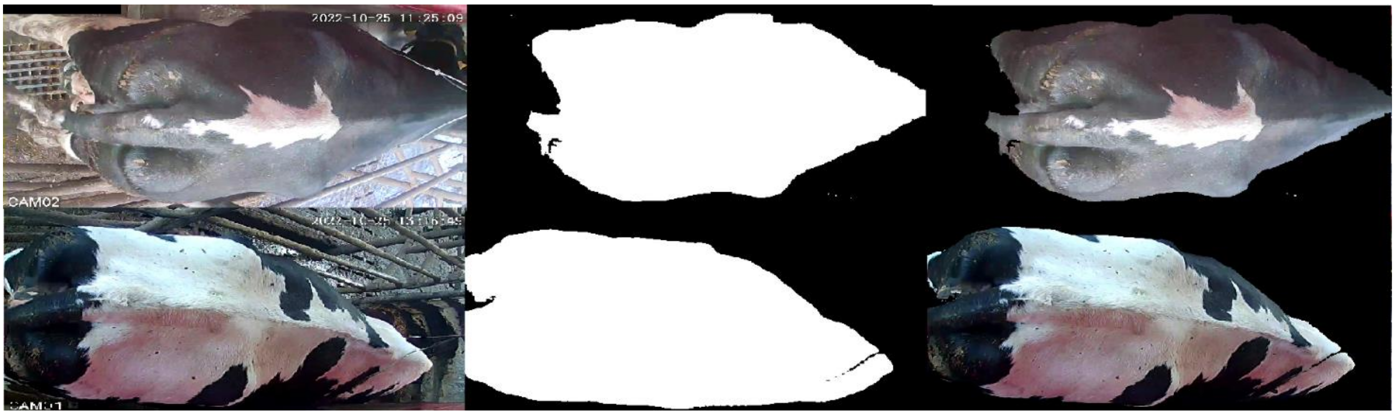

Fig. 3:

Background removal pipeline step-by-step process, on the left the original image, in the middle the mask and on the right the final image.

To address this issue, a series of strategic steps were meticulously executed, leveraging the resources of Google Collab. These measures were undertaken to tackle the challenge of enhancing the model's accuracy and effectively isolating cows within images while transitioning to a cloud-based approach and adhering to TensorFlow's directory structure requirements. Initially, the DONE dataset was uploaded to Google Drive, then the following procedures were carried out:

Segment Anything Model (SAM) Installation

We installed the Segment Anything Model (SAM) developed by Meta AI (Kirillov et al., 2023). SAM is a powerful computer vision model designed for precise object segmentation within images. It excels in identifying and separating objects or regions of interest, regardless of their complexity or variety.

GroundingDINO Integration

GroundingDINO (Liu et al., 2023), another key component of our methodology, was incorporated into the workflow. GroundingDINO is an advanced object detection model that combines elements from Transformer-based detectors and grounded pre-training. It possesses the capability to detect arbitrary objects based on human inputs such as category names or referring expressions, adding an extra layer of flexibility to our approach.

Object Detection and Bounding Box Creation

With GroundingDINO in place, we initiated the process of cow detection. The model effectively identified the cows within the images and generated bounding boxes around them. These bounding boxes outlined the regions containing the cows.

Segmentation with SAM

The Segment Anything Model (SAM)(Kirillov et al., 2023) was then employed to perform precise segmentation. SAM focused exclusively on the area defined by the bounding boxes, effectively creating masks. These masks isolated the cows from the background, producing a clean and distinct representation of each cow within the images.

Mask Application

Finally, the masks generated by SAM were overlaid on the original images. This step effectively removed all elements except for the cows, resulting in images featuring only the isolated cow subjects.

This refined approach allowed us to enhance the accuracy of our model by ensuring that it was no longer trying to recognize the entire image but instead focusing on the crucial task of detecting and recognizing individual cows. By strategically combining the capabilities of SAM and GroundingDINO, we achieved a more precise and efficient method for isolating cows within images, contributing to the overall success of our study (Fig. 3). Thereafter, the edited dataset was moved to the Azure blob.

2.4 Retraining and Integration of cloud-based approach

In the subsequent project phase, our aim was to seamlessly adapt our algorithm to a cloud-based infrastructure, reducing its reliance on local storage resources. This strategic shift was prompted by our commitment to enhance scalability and accessibility. However, this transition towards a cloud-centric approach proved to be more intricate than initially expected, primarily due to the notable incompatibility between specific Tensorflow(Martín~Abadi et al., 2015) classes and the Azure file system. While we addressed all six essential functions and successfully implemented the first five, the last function encountered challenges related to file uploading due to compatibility issues with the Azure file system.

As of today, our code has not achieved full adaptability to the cloud environment. Although the model functions, minor implementations are required for complete compatibility with the cloud-based architecture. These necessities arose due to unexpected challenges stemming from the Tensorflow-Azure interaction. Extensive investigation and dedicated problem-solving efforts have yielded valuable insights into the necessary coding and system configuration adjustments. The ongoing refinement and fine-tuning of our algorithm play a pivotal role in ensuring resilience, effectiveness, and seamless integration within the chosen cloud environment. These continuous efforts are integral to advancing the overall success and scalability of our project.

2.5 Metrics

In our efforts to evaluate the progress of our model's training, we rely on two pivotal metrics: the silhouette score, an evaluator of clustering performance, and cosine similarity, a measure of vector similarity. These metrics serve as essential tools for continuously assessing our model's performance throughout the training journey.

Silhouette Score

The silhouette score provides us with insights into how effectively an individual embedding (referred to as 'i') has been clustered concerning other embeddings. To compute this score, we commence by determining the average distance from 'i' to all embeddings sharing the same label, which we denote as 'a(i)'. Subsequently, we calculate the average distance from 'i' to all embeddings within the closest cluster, denoted as 'b(i)'. The silhouette score ('s(i)'), expressed by the formula:

$$s(i) := \frac{b(i) - \ a(i)}{max \ (a(i), b(i))}$$

This score spans from -1, indicating suboptimal clustering, to 1, signifying exemplary clustering. By calculating the silhouette score at each training step, we gain invaluable insights into the progress of our model's training (Rousseeuw, 1987).

Cosine Similarity

Cosine similarity is a measure of vector similarity (Jiawei Han, 2012), ranging from -1 to 1. Its definition is as follows:

$$cos(\theta) := \frac{A \ \cdot \ B}{|(|A|)| * \ |(|B|)|}$$

A value of 1 denotes identical vectors, while -1 suggests opposite vectors. Within our context, we harness cosine similarity to predict the label of an unknown embedding from a collection of known embeddings. This prediction hinges on identifying the embedding with the highest cosine similarity to the unknown one. Cosine similarity plays a pivotal role during the inference phase of our SNN, facilitating the determination of a cow's identity in an image based on the similarities between embeddings.

3. Results

3.1 Image Extraction

Through the collaborative utilization of GroundingDINO and SAM, we have effectively implemented a comprehensive pipeline for the precise isolation of cows within our images. This refined methodology seamlessly extends its applicability to the raw images used during the inference phase, enabling accurate cow detection. Subsequently, the model assesses whether the cow is in the act of drinking. If this behavior is detected, our pipeline ensures the removal of background elements, resulting in images that are specifically refined for further analysis within the Siamese Neural Network (SNN). The SNN plays a pivotal role in classifying and distinguishing between individual cows, thereby making a significant contribution to the overall success and accuracy of our project.

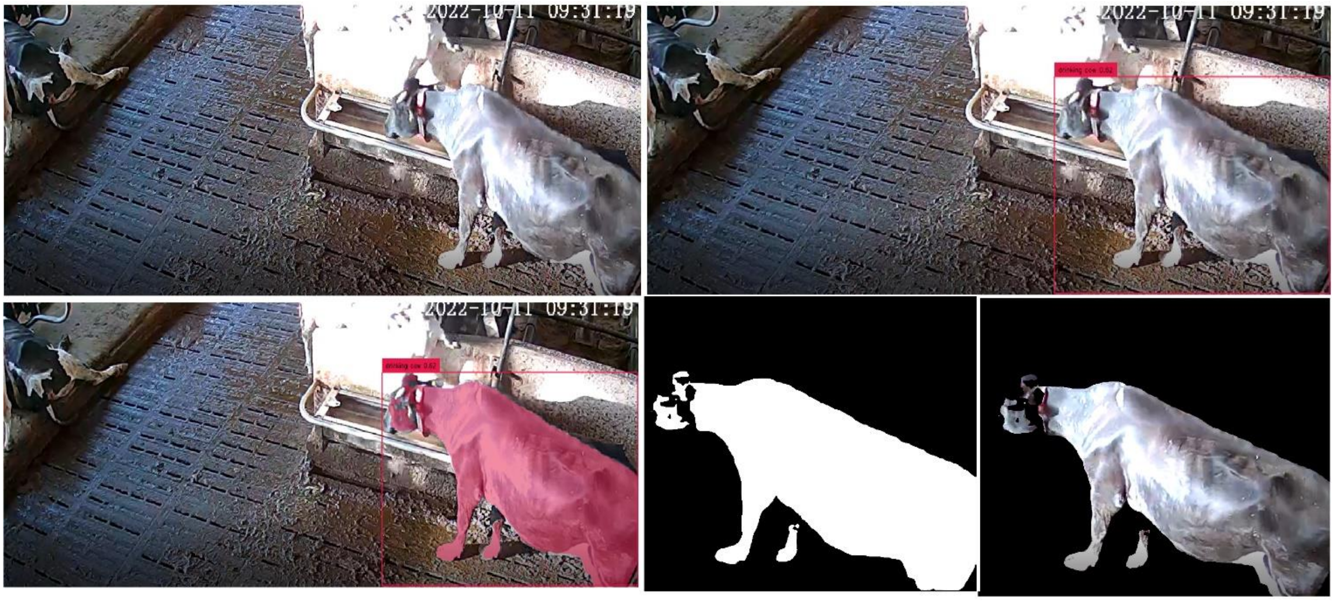

Fig. 4:
Detection of drinking cow and removal of the background

3.2 Model Improvements

After the various implementations and retraining of the model, as explained in sections 2.3 and 2.4, the model achieved its best performance, demonstrating avoidance of overfitting after 50 epochs. As a result, the model checkpoints were extracted and evaluated. This evaluation showed a significant improvement in the Loss score compared to the original model, with a reduction of -0.35 for training (from 0.45 to 0.1) and -0.38 for validation (from 0.5 to 0.12). These improvements signify a substantial decrease in errors and demonstrate the model's enhanced accuracy and precision.

Additionally, the Silhouette scores displayed remarkable progress, with an increase of 0.23 for training (from 0.05 to 0.28) and 0.4 for validation (from 0.1 to 0.5) compared to the previous model. This indicates that the latest model excels in effectively clustering similar cows together. This enhancement reflects the model's ability to consistently and accurately group cows in images, a notable improvement from its previous performance.

Furthermore, the accuracy of the model has seen significant gains during validation, with a 0.31 improvement when provided with an unedited image (from 0.39 to 0.7) and of a 0.76 increase when given an image with the background removed (from 0.1 to 0.86). Overall, these findings indicate that the model has become considerably more robust and is now highly effective in detecting cows with a high degree of certainty. These improvements signify a substantial leap in the model's performance, demonstrating its enhanced accuracy and consistency in identifying and classifying cows in images.

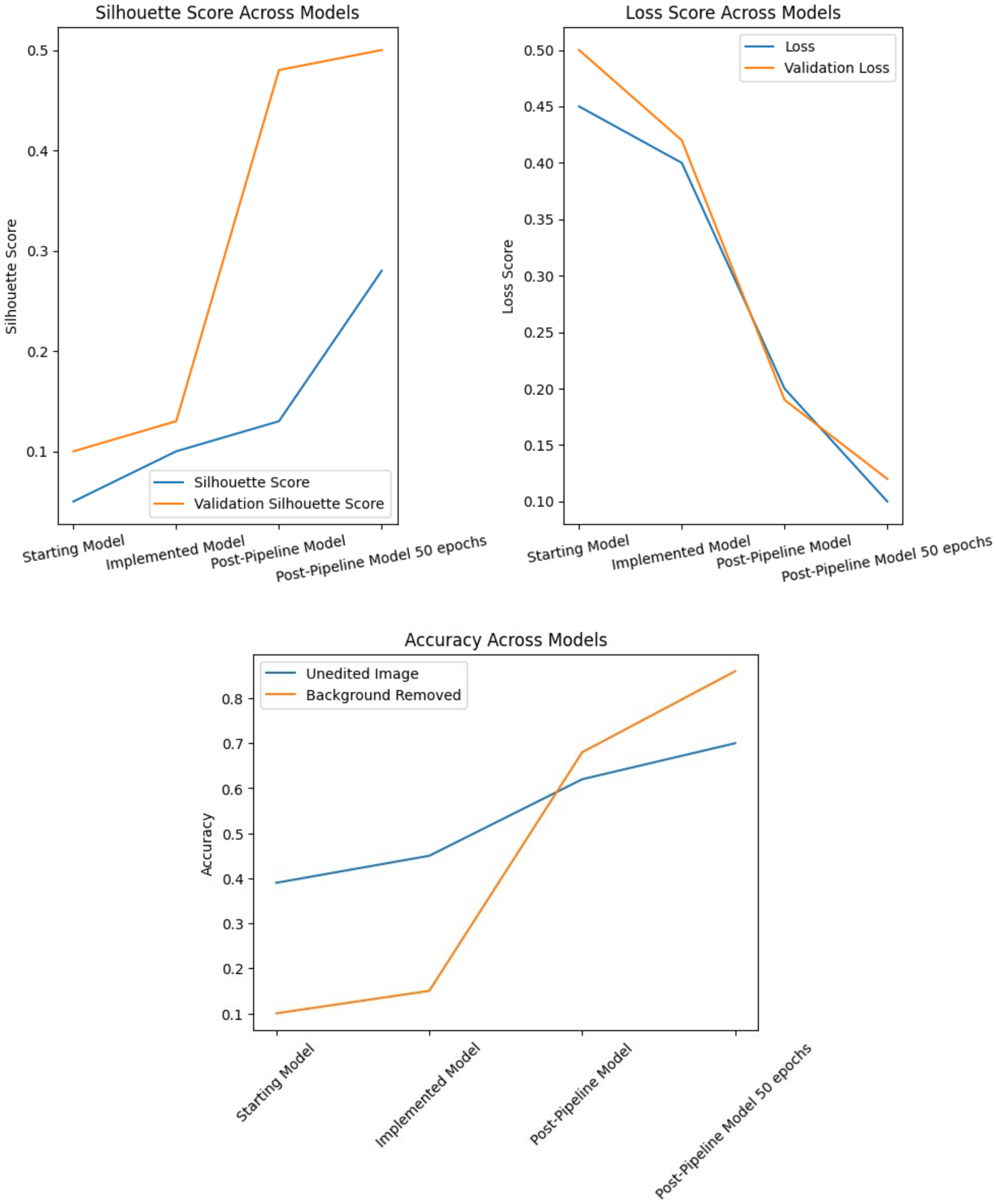

Fig. 5:
Graphic representation of Loss score, Silhouette score and Accuracy across multiple model steps.

3.3 Fairness

Given the scope and time constraints of this substantial project, the primary objective was not to complete it in its entirety but to lay a solid foundation for its continuation. To achieve this, we placed significant emphasis on creating comprehensive documentation and adhering to the FAIR principles (Wilkinson et al., 2016), ensuring that the work remains accessible and extensible. Here's an expanded and improved version of the text:

Code Availability

Extensive measures were taken to ensure the availability and comprehensibility of the codebase. The entire codebase was meticulously documented, employing a systematic approach with clear structure and variable naming that adhered to conventions. In addition, we provided a comprehensive HTML file, produced with sphinx (Brandl, 2021), that explains its functionality and offers guidance on utilizing the major classes integrated into the algorithm. To further assist users and researchers, the code and documentation have been made publicly available on my personal GitHub repository. This repository also includes Jupyter notebooks that exemplify how the code can be effectively applied in various scenarios. Furthermore, to facilitate code execution, a requirements-file is included within the repository, enabling users to effortlessly install all the requisite Python packages.

Data Availability

The collected data, a valuable asset to this project, has been securely stored on an external hard drive entrusted to the corresponding author and has since been returned upon the project's completion. We took great care to meticulously structure and annotate the data, ensuring its traceability and utility. For those interested in accessing the data, it is available upon request, reflecting our commitment to transparency and responsible data management.

4. Discussion

4.1 Future implementations of the model

At its current stage, our model adeptly detects whether a cow is engaged in drinking behavior and can distinguish between individual cows with the necessary cow-specific data. Nevertheless, an exciting avenue for potential future development lies in expanding the model's capabilities to measure the duration of a cow's drinking activity, which is essential for accurately estimating water consumption. Our vision for the next phase includes implementing Grounding Dino to handle video file formats. This enhancement would enable continuous monitoring of cows' drinking patterns and durations, providing a more comprehensive view of their drinking behaviors. With this extended functionality, we can proceed to model the rate at which cows drink, a critical factor for assessing their water consumption. In theory, once these milestones are achieved, the model could effectively determine if each bovine meets its water intake requirements.

This innovative approach involves strategically placing cameras near water troughs, making our model a more cost-effective and versatile alternative to the current RFID (Radio Frequency Identification) monitoring sensors. Such an implementation would represent a significant advancement in our project's capabilities, contributing to more efficient cattle management.

At its current stage, our model demonstrates strong performance. However, an intriguing avenue for exploration involves assessing its adaptability in diverse scenarios, including herds that haven't been previously tested and even within specific cattle breeds that lack distinctive coat motifs. The successful performance of the model under these circumstances would mark a significant achievement.

Envisioning the future, one possibility is to transition our model into a user-friendly application. By doing so, we could facilitate its widespread use across farms worldwide, ensuring its accessibility to diverse agricultural settings. This expansion would represent a remarkable step forward in making our algorithm readily available and beneficial to the global farming community.

4.2 Possible applications of the model

Leveraging the model's ease of trainability and its remarkable performance scores, the algorithm opens the door to numerous prospective applications, particularly when fine-tuned for specific contexts. A captivating avenue for exploration involves broadening the model's utility by enabling it to monitor a diverse spectrum of behaviors among cows. This extension could be realized through the utilization of techniques like Grounding Dino or Deep Lab Cut to model the requisite behaviors, followed by input to the Siamese Neural Network (SNN) for cow detection. Such an enhancement holds the potential to unveil anomalies within the herd, a pivotal aspect of effective livestock management. This exciting possibility harmonizes with our continuous dedication to optimizing the model's capabilities across a range of scenarios.

Barkema, H. W., von Keyserlingk, M. A. G., Kastelic, J. P., Lam, T. J. G. M., Luby, C., Roy, J. P., LeBlanc, S. J., Keefe, G. P., & Kelton, D. F. (2015). Invited review: Changes in the dairy industry affecting dairy cattle health and welfare. *Journal of Dairy Science*, *98*(11), 7426–7445. https://doi.org/10.3168/JDS.2015-9377

Brandl, G. (2021). Sphinx documentation. *URL Http://Sphinx-Doc. Org/Sphinx. Pdf*.

Cardot, V., Le Roux, Y., & Jurjanz, S. (2008). Drinking behavior of lactating dairy cows and prediction of their water intake. *Journal of Dairy Science*, *91*(6), 2257–2264. https://doi.org/10.3168/jds.2007-0204

Henning Steinfeld, P. G. T. D. W. F. and A. O. of the U. N. V. C. C. de H. (2006). *Livestock's Long Shadow: Environmental Issues and Options* (Vol. 24).

Houpt, T. (1984). *Water balance and excretion* (S. M.J, Ed.; 10th ed.). Cornstock Publishing Co.

How much milk does a cow produce? (2023). https://longreads.cbs.nl/the-netherlands-in-numbers-2023/how-much-milk-does-a-cow-produce/

Huang, T. S. (1996). *Computer Vision : Evolution And Promise*. https://doi.org/10.5170/CERN-1996-008.21

Huuskonen, A., Tuomisto, L., & Kauppinen, R. (2011). Effect of drinking water temperature on water intake and performance of dairy calves. *Journal of Dairy Science*, *94*(5), 2475–2480. https://doi.org/10.3168/jds.2010-3723

Jensen, M. B., & Vestergaard, M. (2021). Invited review: Freedom from thirst—Do dairy cows and calves have sufficient access to drinking water? *Journal of Dairy Science*, *104*(11), 11368–11385. https://doi.org/10.3168/JDS.2021-20487

Jiawei Han, M. K. J. P. (2012). *Data Mining*. Elsevier. https://doi.org/10.1016/C2009-0-61819-5

Kertz, A. F., Reutzel, L. F., & Mahoney, J. H. (n.d.). Ad Libitum Water Intake by Neonatal Calves and Its Relationship to Calf Starter Intake, Weight Gain, Feces Score, and Season. *Journal of Dairy Science*, *67*, 2964–2969. https://doi.org/10.3168/jds.S0022-0302(84)81660-4

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

Koch, G., Zemel, R., & Salakhutdinov, R. (n.d.). *Siamese Neural Networks for One-shot Image Recognition*.

Li, S., Fu, L., Sun, Y., Mu, Y., Chen, L., Li, J., & Gong, H. (2021). Individual dairy cow identification based on lightweight convolutional neural network. *PLOS ONE*, *16*(11), e0260510. https://doi.org/10.1371/JOURNAL.PONE.0260510

Liang, Y., Hudson, R. E., & Ballou, M. A. (2020). Supplementing neonatal Jersey calves with a blend of probiotic bacteria improves the pathophysiological response to an oral Salmonella enterica serotype Typhimurium challenge. *Journal of Dairy Science*, *103*(8), 7351–7363. https://doi.org/10.3168/jds.2019-17480

Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., Li, C., Yang, J., Su, H., Zhu, J., & Zhang, L. (2023). *Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection*. https://arxiv.org/abs/2303.05499v4

Martín~Abadi, Ashish~Agarwal, Paul~Barham, Eugene~Brevdo, Zhifeng~Chen, Craig~Citro, Greg~S.~Corrado, Andy~Davis, Jeffrey~Dean, Matthieu~Devin, Sanjay~Ghemawat, Ian~Goodfellow, Andrew~Harp, Geoffrey~Irving, Michael~Isard, Jia, Y., Rafal~Jozefowicz, Lukasz~Kaiser, Manjunath~Kudlur, … Xiaoqiang~Zheng. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. https://www.tensorflow.org/

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience 2018 21:9*, *21*(9), 1281–1289. https://doi.org/10.1038/s41593-018-0209-y

Matthews, S. G., Miller, A. L., Clapp, J., Plötz, T., & Kyriazakis, I. (2016). Early detection of health and welfare compromises through automated detection of behavioural changes in pigs. *The Veterinary Journal*, *217*, 43–51. https://doi.org/10.1016/J.TVJL.2016.09.005

Melin, M., Wiktorsson, H., & Norell, L. (2005). Analysis of feeding and drinking patterns of dairy cows in two cow traffic situations in automatic milking systems. *Journal of Dairy Science*, *88*(1), 71–85. https://doi.org/10.3168/jds.S0022-0302(05)72664-3

Mendes, E. D. M., Carstens, G. E., Tedeschi, L. O., Pinchak, W. E., & Friend, T. H. (2011). Validation of a system for monitoring feeding behavior in beef cattle. *Journal of Animal Science*, *89*(9), 2904–2910. https://doi.org/10.2527/JAS.2010-3489

Meyer, U., Everinghoff, M., Gädeken, D., & Flachowsky, G. (2004). Investigations on the water intake of lactating dairy cows. *Livestock Production Science*, *90*(2–3), 117–121. https://doi.org/10.1016/J.LIVPRODSCI.2004.03.005

Murphy, M. R. (n.d.). *SYMPOSIUM: NUTRITIONAL FACTORS AFFECTING ANIMAL WATER AND WASTE QUALITY Water Metabolism of Dairy Cattle*. https://doi.org/10.3168/jds.S0022-0302(92)77768-6

Nath, T., Mathis, A., Chen, A. C., Patel, A., Bethge, M., & Mathis, M. W. (2019). Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nature Protocols 2019 14:7*, *14*(7), 2152–2176. https://doi.org/10.1038/s41596-019-0176-0

Ridoutt, B. G., Williams, S. R. O., Baud, S., Fraval, S., & Marks, N. (2010). Short communication: The water footprint of dairy products: Case study involving skim milk powder. *Journal of Dairy Science*, *93*(11), 5114–5117. https://doi.org/10.3168/jds.2010-3546

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*(C), 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 815–823. https://doi.org/10.1109/CVPR.2015.7298682

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2016). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *31st AAAI Conference on Artificial Intelligence, AAAI 2017*, 4278–4284. https://doi.org/10.1609/aaai.v31i1.11231

Wang, J., song, Y., Leung, T., Rosenberg, C., Wang, J., Philbin, J., Chen, B., & Wu, Y. (2014). *Learning Fine-grained Image Similarity with Deep Ranking*.

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data 2016 3:1*, *3*(1), 1–9. https://doi.org/10.1038/sdata.2016.18

Zehetmeier, M., Baudracco, J., Hoffmann, H., & Heißenhuber, A. (2012). Does increasing milk yield per cow reduce greenhouse gas emissions? A system approach. *Animal*, *6*(1), 154–166. https://doi.org/10.1017/S1751731111001467